Linear Algebra and Its Applications, Second Edition

Solution of Exercise Problems

Ai Shu Xue

January 26, 2010

Contents

This is a solution manual for the textbook *Linear Algebra and Its Applications*, 2nd Edition, by Peter Lax (John Wiley & Sons, 2007). This version omits the following problems: Exercise 2, 9 of Chapter 8; Exercise 3 of Appendix 3; Exercise problems of Appendix 4, 5, 8 and 11.

1 Fundamentals

1.

Proof. Suppose 0 and $0'$ are two zeros of vector addition, then by the definition of zero and commutativity, we have $0' = 0' + 0 = 0 + 0' = 0$. $\square$

2.

Proof. For any $x = (x_1, \cdots, x_n) \in K^n$, we have

$$x + 0 = (x_1, \cdots, x_n) + (0, \cdots, 0) = (x_1 + 0, \cdots, x_n + 0) = (x_1, \cdots, x_n) = x.$$

So $0 = (0, \cdots, 0)$ is zero element of classical vector addition. $\square$

3.

Proof. The isomorphism T can be defined as $T((a_1, \cdots, a_n)) = a_1 + a_2 x + \cdots + a_n x^{n-1}$. $\square$

4.

Proof. Suppose $S = \{s_1, \cdots, s_n\}$. The isomorphism T can be defined as $T(f) = (f(s_1), \cdots, f(s_n))$,$\forall f \in K^S$. $\square$

5.

Proof. For any $p(x) = a_1 + a_2 x + \cdots + a_n x^{n-1}$, we define

$$T(p) = p(x),$$

where p on the left side of the equation is regarded as a polynomial over $\mathbb{R}$ while $p(x)$ on the right side of the equation is regarded as a function defined on $S = \{s_1, \cdots, s_n\}$. To prove T is an isomorphism, it suffices to prove T is one-to-one. This is seen through the observation that

$$\begin{bmatrix} 1 & s_1 & s_1^2 & \cdots & s_1^{n-1} \\ 1 & s_2 & s_2^2 & \cdots & s_2^{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & s_n & s_n^2 & \cdots & s_n^{n-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} p(s_1) \\ p(s_2) \\ \cdots \\ p(s_n) \end{bmatrix}$$

and the Vandermonde matrix

$$\begin{bmatrix} 1 & s_1 & s_1^2 & \cdots & s_1^{n-1} \\ 1 & s_2 & s_2^2 & \cdots & s_2^{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & s_n & s_n^2 & \cdots & s_n^{n-1} \end{bmatrix}$$

is invertible for distinct s_1, s_2, $\cdots$, s_n. $\square$

6.

Proof. For any $y, y' \in Y$, $z, z' \in Z$ and $k \in K$, we have (by commutativity and associative law)

$$\begin{aligned}(y+z)+(y'+z') &= (z+y)+(y'+z') = z+(y+(y'+z')) = z+((y+y')+z') = z+(z'+(y+y')) \\ &= (z+z')+(y+y') = (y+y')+(z+z') \in Y+Z,\end{aligned}$$

and

$$k(y+z) = ky + kz \in Y + Z.$$

So $Y + Z$ is a linear subspace of X if Y and Z are. □

7.

Proof. For any $x_1, x_2 \in Y \cap Z$, since Y and Z are linear subspaces of X, $x_1 + x_2 \in Y$ and $x_1 + x_2 \in Z$. Therefore, $x_1 + x_2 \in Y \cap Z$. For any $k \in K$ and $x \in Y \cap Z$, since Y and Z are linear subspaces of X, $kx \in Y$ and $kx \in Z$. Therefore, $kx \in Y \cap Z$. Combined, we conclude $Y \cap Z$ is a linear subspace of X. □

8.

Proof. By definition of zero vector, $0+0 = 0 \in \{0\}$. For any $k \in K$, $k0 = k(0+0) = k0+k0$. So $k0 = 0 \in \{0\}$. Combined, we can conclude $\{0\}$ is a linear subspace of X. □

9.

Proof. Define $Y = \{k_1x_1 + \cdots + k_jx_j : k_1, \cdots, k_j \in K\}$. Then clearly $x_1 = 1x_1 + 0x_2 + \cdots + 0x_j \in Y$. Similarly, we can show $x_2, \cdots, x_j \in Y$. Since for any $k_1, \cdots, k_j, k_1', \cdots, k_j' \in K$,

$$(k_1x_1 + \cdots + k_jx_j) + (k_1'x_1 + \cdots + k_j'x_j) = (k_1 + k_1')x_1 + \cdots + (k_j + k_j')x_j \in Y$$

and for any $k_1, \cdots, k_j, k \in K$,

$$k(k_1x_1 + \cdots + k_jx_j) = (kk_1)x_1 + \cdots + (kk_j)x_j \in Y,$$

we can conclude Y is a linear subspace of X containing $x_1, \cdots, x_j$. Finally, if Z is any linear subspace of X containing $x_1, \cdots, x_j$, it is clear that $Y \subset Z$ as Z must be closed under scalar multiplication and vector addition. Combined, we have proven Y is the smallest linear subspace of X containing $x_1, \cdots, x_j$. □

10.

Proof. We prove by contradiction. Without loss of generality, assume $x_1 = 0$. Then $1x_1 + 0x_2 + \cdots + 0x_j = 0$. This shows $x_1, \cdots, x_j$ are linearly dependent, a contradiction. So $x_1 \neq 0$. We can similarly prove $x_2, \cdots, x_j \neq 0$. □

11.

Proof. Suppose Y_i has a basis $y_1^i, \cdots, y_{n_i}^i$. Then it suffices to prove $y_1^1, \cdots, y_{n_1}^1, \cdots, y_1^m, \cdots, y_{n_m}^m$ form a basis of X. By definition of direct sum, these vectors span X, so we only need to show they are linearly independent. In fact, if not, then 0 has two distinct representations: $0 = 0 + \cdots + 0$ and $0 = \sum_{i=1}^m (a_1^i y_1^i + \cdots + a_{n_i}^i y_{n_i}^i)$ for some $a_1^1, \cdots, a_{n_1}^1, \cdots, a_1^m, \cdots, a_{n_m}^m$, where not all a_j^i are zero. This is contradictory with the definition of direct sum. So we must have linear independence, which imply $y_1^1, \cdots, y_{n_1}^1, \cdots, y_1^m, \cdots, y_{n_m}^m$ form a basis of X. Consequently, $\dim X = \sum \dim Y_i$. □

12.

Proof. Fix a basis $x_1, \cdots, x_n$ of X, any element $x \in X$ can be uniquely represented as $\sum_{i=1}^n \alpha_i(x)x_i$ for some $\alpha_i(x) \in K$, $i = 1, \cdots, n$. We define the isomorphism as $x \mapsto (\alpha_1(x), \cdots, \alpha_n(x))$. Clearly this isomorphism depends on the basis and by varying the choice of basis, we can have different isomorphisms. □

13.

Proof. For any $x_1, x_2 \in X$, if $x_1 \equiv x_2$, i.e. $x_1 - x_2 \in Y$, then $x_2 - x_1 = -(x_1 - x_2) \in Y$, i.e. $x_2 \equiv x_1$. This is symmetry. For any $x \in X$, $x - x = 0 \in Y$. So $x \equiv x$. This is reflexivity. Finally, if $x_1 \equiv x_2$, $x_2 \equiv x_3$, then $x_1 - x_3 = (x_1 - x_2) + (x_2 - x_3) \in Y$, i.e. $x_1 \equiv x_3$. This is transitivity. □

14.

Proof. For any $x_1, x_2 \in X$, we can find $y \in \{x_1\} \cap \{x_2\}$ if and only if $x_1 - y \in Y$ and $x_2 - y \in Y$. Then

$$x_1 - x_2 = (x_1 - y) - (x_2 - y) \in Y.$$

So $\{x_1\} \cap \{x_2\} \neq \emptyset$ if and only if $\{x_1\} = \{x_2\}$. □

15.

Proof. If $\{x\} = \{x'\}$ and $\{y\} = \{y'\}$, then $x - x', y - y' \in Y$. So $(x+y) - (x'+y') = (x - x') + (y - y') \in Y$. This shows $\{x + y\} = \{x' + y'\}$. Also, for any $k \in K$, $kx - kx' = k(x - x') \in Y$. So $k\{x\} = \{kx\} = \{kx'\} = k\{x'\}$. □

16.

Proof. By theory of polynomials, we have

$$Y = \left\{ q(t) \prod_{i=1}^{j} (t - t_i) : q(t) \text{ is a polynomial of degree } < n - j \right\}.$$

Then it's easy to see $\dim Y = n - j$ and $\dim X/Y = \dim X - \dim Y = j$. □

17.

Proof. By Theorem 6, $\dim X/Y = \dim X - \dim Y = 0$, which implies $X/Y = \{\{0\}\}$. So $X = Y$. □

18.

Proof. Define $Y_1 = \{(x, 0) : x \in X_1, 0 \in X_2\}$ and $Y_2 = \{(0, x) : 0 \in X_1, x \in X_2\}$. Then Y_1 and Y_2 are linear subspaces of $X_1 \oplus X_2$. It is easy to see Y_1 is isomorphic to X_1, Y_2 is isomorphic to X_2, and $Y_1 \cap Y_2 = \{(0, 0)\}$. So by Theorem 7, $\dim X_1 \oplus X_2 = \dim Y_1 + \dim Y_2 - \dim(Y_1 \cap Y_2) = \dim X_1 + \dim X_2 - 0 = \dim X_1 + \dim X_2$. □

19.

Proof. By Exercise 18 and Theorem 6, $\dim(Y \oplus X/Y) = \dim Y + \dim(X/Y) = \dim Y + \dim X - \dim Y = \dim X$. Since linear spaces of same finite dimension are isomorphic (by one-to-one mapping between their bases), $Y \oplus X/Y$ is isomorphic to X. □

20.

Proof. (a) is not since $\{x : x_1 \geq 0\}$ is not closed under the scalar multiplication by -1. (b) is. (c) is not since $x_1 + x_2 + 1 = 0$ and $x_1' + x_2' + 1 = 0$ imply $(x_1 + x_1') + (x_2 + x_2') + 1 = -1$. (d) is. (e) is not since x_1 being an integer does not guarantee rx_1 is an integer for any $r \in \mathbb{R}$. □

21.

Proof. See the textbook's solution, page 279. □

2 Duality

1.

Proof. Suppose $e_1, \cdots, e_n$ is a basis of X and suppose $x_1 = \sum_{i=1}^n \alpha_i e_i$. If the underlying field is $\mathbb{R}$, we define a linear function l by setting $l(e_i) = a_i$ $(i = 1, \cdots, n)$ and extending its definition to X via linear combination. If the underlying field is $\mathbb{C}$, we define l similarly by setting $l(e_i) = a_i^*$, where a_i^* is the complex conjugate of a_i $(i = 1, \cdots, n)$. In either case, we have $l(x_1) = \|x_1\|^2 \neq 0$, where $\|\cdot\|$ is the Euclidean norm of $\mathbb{R}^n$ or $\mathbb{C}^n$.

To generalize the above result to a general linear vector space X over a field K, we clearly need some notion of *norm*. This is exactly the starting point of Hahn-Banach Theorem, which claims a similar result for general linear vector spaces, not necessarily finite-dimensional (see Lax [6]). So this exercise problem needs extra conditions if we need to go beyond $K = \mathbb{R}$ or $K = \mathbb{C}$. $\square$

2.

Proof. For any l_1 and $l_2 \in Y^\perp$, we have $(l_1+l_2)(y) = l_1(y)+l_2(y) = 0+0 = 0$ for any $y \in Y$. So $l_1+l_2 \in Y^\perp$. For any $k \in K$, $(kl)(y) = k(l(y)) = k0 = 0$ for any $y \in Y$. So $kl \in Y^\perp$. Combined, we conclude $Y^\perp$ is a subspace of X'. $\square$

3.

Proof. Since $S \subset Y$, $Y^\perp \subset S^\perp$. For "$\supset$", let $x_1, \cdots, x_m$ be a maximal linearly independent subset of S. Then $S = \text{span}(x_1, \cdots, x_m)$ and $Y = \{\sum_{i=1}^m \alpha_i x_i : \alpha_1, \cdots, \alpha_m \in K\}$ by Exercise 9 of Chapter 1. By the definition of annihilator, for any $l \in S^\perp$ and $y = \sum_{i=1}^m \alpha_i x_i \in Y$, we have

$$l(y) = \sum_{i=1}^m \alpha_i l(x_i) = 0.$$

So $l \in Y^\perp$. By the arbitrariness of l, $S^\perp \subset Y^\perp$. Combined, we have $S^\perp = Y^\perp$. $\square$

4.

Proof. Suppose three linearly independent polynomials p_1, p_2 and p_3 are applied to formula (9). Then m_1, m_2 and m_3 must satisfy the linear equations

$$\begin{bmatrix} p_1(t_1) & p_1(t_2) & p_1(t_3) \\ p_2(t_1) & p_2(t_2) & p_2(t_3) \\ p_3(t_1) & p_3(t_2) & p_3(t_3) \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} \int_{-1}^1 p_1(t)dt \\ \int_{-1}^1 p_2(t)dt \\ \int_{-1}^1 p_3(t)dt \end{bmatrix}$$

We take $p_1(t) = 1$, $p_2(t) = t$ and $p_3(t) = t^2$. The above equation becomes

$$\begin{bmatrix} 1 & 1 & 1 \\ -a & 0 & a \\ a^2 & 0 & a^2 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ \frac{2}{3} \end{bmatrix}$$

So

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -a & 0 & a \\ a^2 & 0 & a^2 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2a} & \frac{1}{2a^2} \\ 1 & 0 & -\frac{1}{a^2} \\ 0 & \frac{1}{2a} & \frac{1}{2a^2} \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{3a^2} \\ 2 - \frac{2}{3a^2} \\ \frac{1}{3a^2} \end{bmatrix}$$

Then it's easy to see that for $a > \sqrt{1/3}$, all three weights are positive.

To show formula (9) holds for all polynomials of degree < 6 when $a = \sqrt{3/5}$, we note for any odd $n \in \mathbb{N}$,

$$\int_{-1}^1 x^n dx = 0, \ m_1 p(-a) + m_3 p(a) = 0 \text{ since } m_1 = m_2 \text{ and } p(-x) = -p(x), \text{ and } m_2 p(0) = 0.$$

So (9) holds for any x^n of odd degree n. In particular, for $p(x) = x^3$ and $p(x) = x^5$. For $p(x) = x^4$, we have

$$\int_{-1}^{1} x^4 dx = \frac{2}{5},\ m_1 p(t_1) + m_2 p(t_2) + m_3 p(t_3) = 2m_1 a^4 = \frac{2}{3}a^2.$$

So formula (9) holds for $p(x) = x^4$ when $a = \sqrt{3/5}$. Combined, we conclude for $a = \sqrt{3/5}$, (9) holds for all polynomials of degree < 6. $\square$

Remark 1. *In this exercise problem and Exercise 5 below, "Theorem 6" should be corrected to "Theorem 7".*

5.

Proof. We take $p_1(t) = 1$, $p_2(t) = t$, $p_3(t) = t^2$, and $p_4(t) = t^3$. Then m_1, m_2, m_3, and m_4 solve the following equation:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -a & -b & b & a \\ a^2 & b^2 & b^2 & a^2 \\ -a^3 & -b^3 & b^3 & a^3 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2/3 \\ 0 \end{bmatrix}$$

Then

$$\begin{aligned} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ -a & -b & b & a \\ a^2 & b^2 & b^2 & a^2 \\ -a^3 & -b^3 & b^3 & a^3 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 0 \\ 2/3 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{b^2}{-2a^2+2b^2} & \frac{b^2}{2a^3-2ab^2} & \frac{1}{2a^2-2b^2} & \frac{1}{-2a^3+2ab^2} \\ \frac{a^2}{2a^2-2b^2} & \frac{a^2}{-2a^2b+2b^3} & \frac{1}{-2a^2+2b^2} & \frac{1}{2a^2b-2b^3} \\ \frac{a^2}{2a^2-2b^2} & \frac{a^2}{2a^2b-2b^3} & \frac{1}{-2a^2+2b^2} & \frac{1}{-2a^2b+2b^3} \\ \frac{b^2}{-2a^2+2b^2} & \frac{b^2}{-2a^3+2ab^2} & \frac{1}{2a^2-2b^2} & \frac{1}{2a^3-2ab^2} \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 2/3 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{-3b^2+1}{3(a^2-b^2)} \\ \frac{3a^2-1}{3(a^2-b^2)} \\ \frac{3a^2-1}{3(a^2-b^2)} \\ \frac{-3b^2+1}{3(a^2-b^2)} \end{bmatrix} \end{aligned}$$

So the weights are positive if and only if one of the following two mutually exclusive cases hold
1) $b^2 > \frac{1}{3}$, $a^2 < b^2$, $a^2 > \frac{1}{3}$;
2) $b^2 < \frac{1}{3}$, $a^2 > b^2$, $a^2 < \frac{1}{3}$.

$\square$

6.

Proof. (from the textbook's solution) (a) Suppose there is a linear relation

$$al_1(p) + bl_2(p) + cl_3(p) = 0.$$

Set $p = p(x) = (x - \xi_2)(x - \xi_3)$. Then $p(\xi_2) = p(\xi_3) = 0$, $p_1(\xi_1) \neq 0$; so we get from the above relation that $a = 0$. Similarly $b = 0$, $c = 0$.

(b) Since $\dim P_2 = 3$, also $\dim P_2' = 3$. Since l_1, l_2, l_3 are linearly independent, the span P_2'.

(c1) We define l_1 by setting

$$l_1(e_j) = \begin{cases} 1, & \text{if } j = 1 \\ 0, & \text{if } j \neq 1 \end{cases}$$

and extending l_1 to V by linear combination, i.e. $l_1(\sum_{j=1}^n \alpha_j e_j) := \sum_{j=1}^n \alpha_j l_1(e_j) = \alpha_1$. $l_2, \cdots, l_n$ can be constructed similarly. If there exist $a_1, \cdots, a_n$ such that $a_1 l_1 + \cdots + a_n l_n = 0$, we have

$$0 = a_1 l_1(e_j) + \cdots a_n l_n(e_j) = a_j,\ j = 1, \cdots, n.$$

So $l_1, \cdots, l_n$ are linearly independent. Since $\dim V' = \dim V = n$, $\{l_1, \cdots, l_n\}$ is a basis of V'.

(c2) We define

$$p_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)},\ p_2(x) = \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)},\ p_3(x) = \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}.$$

□

7.

Proof. (from the textbook's solution) $l(x)$ has to be zero for $x = (1, 0, -1, 2)$ and $x = (2, 3, 1, 1)$. These yield two equations for $c_1, \cdots, c_4$:

$$c_1 - c_3 + 2c_4 = 0,\ 2c_1 + 3c_2 + c_3 + c_4 = 0.$$

We express c_1 and c_2 in terms of c_3 and c_4. From the first equation, $c_1 = c_3 - 2c_4$. Setting this into the second equation gives $c_2 = -c_3 + c_4$. □

3 Linear Mappings

1. (a)

Proof. For any $y, y' \in T(X)$, there exist $x, x' \in X$ such that $T(x) = y$ and $T(x') = y'$. So $y + y' = T(x) + T(x') = T(x + x') \in T(X)$. For any $k \in K$, $ky = kT(x) = T(kx) \in T(X)$. Combined, we conclude $T(X)$ is a linear subspace of U. □

(b)

Proof. Suppose V is a linear subspace of U. For any $x, x' \in T^{-1}(V)$, there exist $y, y' \in V$ such that $T(x) = y$ and $T(x') = y'$. Since $T(x + x') = T(x) + T(x') = y + y' \in V$, $x + x' \in T^{-1}(V)$. For any $k \in K$, since $T(kx) = kT(x) = ky \in V$, $kx \in T^{-1}(V)$. Combined, we conclude $T^{-1}(V)$ is a linear subspace of X. □

2.

Proof. (from the textbook's solution) Suppose we drop the ith equation; if the remaining equations do not determine x uniquely, there is an x that is mapped into a vector whose components except the ith are zero. If this were true for all $i = 1, \cdots, m$, the range of the mapping $x \to u$ would be m-dimensional; but according to Theorem 2, the dimension of the range is $\leq n < m$. Therefore one of the equations may be dropped without losing uniqueness; by induction $m - n$ of the equations may be omitted.

Alternative solution: Uniqueness of the solution x implies the column vectors of the matrix $T = (t_{ij})$ are linearly independent. Since the column rank of a matrix equals its row rank (see Chapter 4), it is possible to select a subset of n of these equations which uniquely determine the solution. □

Remark 2. *The textbook's solution is a proof that the column rank of a matrix equals its row rank.*

3. (i)

Proof. $S \circ T(ax + by) = S(T(ax + by)) = S(aT(x) + bT(y)) = aS(T(x)) + bS(T(y)) = aS \circ T(x) + bS \circ T(y)$. So $S \circ T$ is also a linear mapping. □

(ii)

Proof. $(R + S) \circ T(x) = (R + S)(T(x)) = R(T(x)) + S(T(x)) = (R \circ T + S \circ T)(x)$ and $S \circ (T + P)(x) = S((T + P)(x)) = S(T(x) + P(x)) = S(T(x)) + S(P(x)) = (S \circ T + S \circ P)(x)$. □

4. (a)

Proof. Linearity of S and T is easy to see. For non-commutativity, consider the polynomial s. Then $TS(s) = T(s^2) = 2s \neq s = S(1) = ST(s)$. So $ST \neq TS$. □

(b)

Proof. For any $x = (x_1, x_2, x_3) \in X$, $S(x) = (x_1, x_3, -x_2)$ and $T(x) = (x_3, x_2, -x_1)$. So it's easy to see S and T are linear. For non-commutativity, note $ST(x) = S(x_3, x_2, -x_1) = (x_3, -x_1, -x_2)$ and $TS(x) = T(x_1, x_3, -x_2) = (-x_2, x_3, -x_1)$. So $ST \neq TS$ in general. □

Remark 3. *Note the problem does not specify the direction of the rotation, so it is also possible that $S(x) = (x_1, -x_3, x_2)$ and $T(x) = (-x_3, x_2, x_1)$. There are total of four choices of (S, T). But the corresponding proofs are similar to the one presented here.*

5.

Proof. $TT^{-1}(x) = T(T^{-1}(x)) = x$ by definition. So $TT^{-1} = id$. □

6. (i)

Proof. Suppose $T : X \to U$ is invertible. Then for any $y, y' \in U$, there exist a unique $x \in X$ and a unique $x' \in X$ such that $T(x) = y$ and $T(x') = y'$. So $T(x + x') = T(x) + T(x') = y + y'$ and by the injectivity of T, $T^{-1}(y + y') = x + x' = T^{-1}(y) + T^{-1}(y')$. For any $k \in K$, since $T(kx) = kT(x) = ky$, injectivity of T implies $T^{-1}(ky) = kx = kT^{-1}(y)$. Combined, we conclude T^{-1} is linear. □

(ii)

Proof. Suppose $T : X \to U$ and $S : U \to V$. First, by the definition of multiplication, ST is a linear map. Second, if $x \in X$ is such that $ST(x) = 0 \in V$, the injectivity of S implies $T(x) = 0 \in U$ and the injectivity of T further implies $x = 0 \in X$. So, ST is one-to-one. For any $z \in V$, there exists $y \in U$ such that $S(y) = z$. Also, we can find $x \in X$ such that $T(x) = y$. So $ST(x) = S(y) = z$. This shows ST is onto. Combined, we conclude ST is invertible.

By associativity, we have $(ST)(T^{-1}S^{-1}) = ((ST)T^{-1})S^{-1} = (S(TT^{-1}))S^{-1} = SS^{-1} = \mathrm{id}_V$. Replace S with T^{-1} and T with S^{-1}, we also have $(T^{-1}S^{-1})(ST) = \mathrm{id}_X$. Therefore, we can conclude $(ST)^{-1} = T^{-1}S^{-1}$. □

7. (i)

Proof. Suppose $T : X \to U$ and $S : U \to V$ are linear maps. Then for any given $l \in V'$, $((ST)'l, x) = (l, STx) = (S'l, Tx) = (T'S'l, x)$, $\forall x \in X$. Therefore, $(ST)'l = T'S'l$. Let l run through every element of V', we conclude $(ST)' = T'S'$. □

(ii)

Proof. Suppose T and R are both linear maps from X to U. For any given $l \in U'$, we have $((T + R)'l, x) = (l, (T + R)x) = (l, Tx + Rx) = (l, Tx) + (l, Rx) = (T'l, x) + (R'l, x) = ((T' + R')l, x)$, $\forall x \in X$. Therefore $(T + R)'l = (T' + R')l$. Let l run through every element of V', we conclude $(T + R)' = T' + R'$. □

(iii)

Proof. Suppose T is an isomorphism from X to U, then T^{-1} is a well-defined linear map. We first show T' is an isomorphism from U' to X'. Indeed, if $l \in U'$ is such that $T'l = 0$, then for any $x \in X$, $0 = (T'l, x) = (l, Tx)$. As x varies and goes through every element of X, Tx goes through every element of U. By considering the identification of U with U'', we conclude $l = 0$. So T' is one-to-one. For any given $m \in X'$, define $l = mT^{-1}$, then $l \in U'$. For any $x \in X$, we have $(m, x) = (m, T^{-1}(Tx)) = (l, Tx) = (T'l, x)$. Since x is arbitrary, $m = T'l$ and T' is therefore onto. Combined, we conclude T' is an isomorphism from U' to X' and $(T')^{-1}$ is hence well-defined.

By part (i), $(T^{-1})'T' = (TT^{-1})' = (\mathrm{id}_U)' = \mathrm{id}_{U'}$ and $T'(T^{-1})' = (T^{-1}T)' = (\mathrm{id}_X)' = \mathrm{id}_{X'}$. This shows $(T^{-1})' = (T')^{-1}$. □

8.

Proof. Suppose $\xi : X \to X''$ and $\eta : U \to U''$ are the isomorphisms defined in Chapter 2, formula (5), which identify X with X'' and U with U'', respectively. Then for any $x \in X$ and $l \in U'$, we have

$$(T''\xi_x, l) = (\xi_x, T'l) = (T'l, x) = (l, Tx) = (\eta_{Tx}, l).$$

Since l is arbitrary, we must have $T''\xi_x = \eta_{Tx}$, $\forall x \in X$. Hence, $T'' \circ \xi = \eta \circ T$, which is the precise interpretation of $T'' = T$. □

9.

Proof. If $Bx = 0$, by applying A to both sides of the equation and $AB = I$, we conclude $x = 0$. So B is injective. By Corollary B of Theorem 2, B is surjective. Therefore the inverse of B, denoted by B^{-1}, always exists, and $A = A(BB^{-1}) = (AB)B^{-1} = IB^{-1} = B^{-1}$, which implies $BA = I$. □

Remark 4. *For a general algebraic structure, e.g. a ring with unit, it's not always the case that an element's right inverse equals to its left inverse. In the proof above, we used the fact that for finite dimensional linear vector space, a linear mapping is injective if and only if it's surjective.*

10.

Proof. Suppose $K = M_S$. Then $K(M^{-1})_S = SMS^{-1}SM^{-1}S^{-1} = I$. By Exercise 9, K is also invertible and $K^{-1} = (M^{-1})_S$. □

11.

Proof. Suppose A is invertible, we have $AB = AB(AA^{-1}) = A(BA)A^{-1}$. So AB and BA are similar. The case of B being invertible can be proved similarly. □

12.

Proof. For any $\alpha, \beta \in K$ and $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$, we have

$$\begin{aligned} P(\alpha x + \beta y) &= P((\alpha x_1 + \beta y_1, \cdots, \alpha x_n + \beta y_n)) \\ &= (0, 0, \alpha x_3 + \beta y_3, \cdots, \alpha x_n + \beta y_n) \\ &= (0, 0, \alpha x_3, \cdots, \alpha x_n) + (0, 0, \beta y_3, \cdots, \beta y_n) \\ &= \alpha(0, 0, x_3, \cdots, x_n) + \beta(0, 0, y_3, \cdots, y_n) \\ &= \alpha P(x) + \beta P(y). \end{aligned}$$

This shows P is a linear map. Furthermore, we have

$$P^2(x) = P((0, 0, x_3, \cdots, x_n)) = (0, 0, x_3, \cdots, x_n) = P(x).$$

So P is a projection. □

13.

Proof. For any $\alpha, \beta \in K$ and $f, g \in C[-1, 1]$, we have

$$P(\alpha f + \beta g)(x) = \frac{1}{2}[(\alpha f + \beta g)(x) + (\alpha f + \beta g)(-x)] = \frac{\alpha}{2}[f(x) + f(-x)] + \frac{\beta}{2}[g(x) + g(-x)] = \alpha P(f)(x) + \beta P(g)(x).$$

This shows P is a linear map. Furthermore, we have

$$\begin{aligned} (P^2 f)(x) &= (P(Pf))(x) = P\left(\frac{f(\cdot) + f(-\cdot)}{2}\right)(x) = \frac{1}{2}\left[\frac{f(x) + f(-x)}{2} + \frac{f(-x) + f(x)}{2}\right] \\ &= \frac{1}{2}(f(x) + f(-x)) = (Pf)(x). \end{aligned}$$

So P is a projection. □

14. (a)

Proof. Since $\dim R_T = 1$, it suffices to prove the following claim: *if T is a linear map on a 1-dimensional linear vector space X, there exists a unique number c such that $T(x) = cx$, $\forall x \in X$.* We assume the underlying filed K is either $\mathbb{R}$ or $\mathbb{C}$. We further assume $S : X \to K$ is an isomorphism. Then $S \circ T \circ S^{-1}$ is a linear map on K. Define $c = S \circ T \circ S^{-1}(1)$, we have

$$S \circ T \circ S^{-1}(k) = S \circ T \circ S^{-1}(k \cdot 1) = k \cdot c, \ \forall k \in K.$$

So $T \circ S^{-1}(k) = S^{-1}(c \cdot k) = cS^{-1}(k)$, $\forall k \in K$. This shows T is a scalar multiplication. □

(b)

Proof. If $c \neq 1$, it's easy to verify $I + \frac{1}{1-c}T$ is the inverse of $I - T$. □

15.

Proof. Because $R_{ST} \subset R_S$, $\dim(R_{ST}) \leq \dim R_S$. Also, we note N_{ST}/N_T is isomorphic to $N_S \cap R_T$, with the isomorphism defined by $\phi(\{x\}) = Tx$, where $\{x\} := x + N_T$. It's easy to see ϕ is well-defined, is linear, and is both injective and surjective. So by Theorem 6 of Chapter 1,

$$\dim N_{ST} = \dim N_T + \dim N_{ST}/N_T = \dim N_T + \dim(N_S \cap R_T) \leq \dim N_T + \dim N_S.$$

□

4 Matrices

1.

Proof. It looks the phrasement of the exercise has a problem: when $m \neq n$, AD or DA may not be well-defined. So we will assume $m = n$ in the below. We can write A in the row form $\begin{bmatrix} r_1 \\ r_2 \\ \cdots \\ r_m \end{bmatrix}$. Then DA can be written as

$$DA = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \cdots \\ r_n \end{bmatrix} = \begin{bmatrix} d_1 r_1 \\ d_2 r_2 \\ \cdots \\ d_n r_n \end{bmatrix}$$

We can also write A in the column form $[c_1, c_2, \cdots, c_n]$, then AD can be written as

$$AD = [c_1, c_2, \cdots, c_n] \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} = [d_1 c_1, d_2 c_2, \cdots, d_n c_n]$$

□

2.

Proof. Proofs in most textbooks are lengthy and complicated. For a clear, although still lengthy, proof, see Qiu [10], Theorem 3.5.3, page 112. □

3.

Proof. The calculation is a bit messy. We refer the reader to Qiu [10], Theorem 4.6.1, page 190. □

4.

Proof. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}$. Then $AB = 0$ yet $BA = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \neq 0$. ☐

5.

Proof.

$$\begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & 5 & 4 & -3 \\ 2 & 3 & 4 & 1 \\ 1 & 4 & 2 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 2 + 3 \cdot (-2) + (-1) \cdot 1 \\ 2 \cdot 1 + 5 \cdot 2 + 4 \cdot (-2) + (-3) \cdot 1 \\ 2 \cdot 1 + 3 \cdot 2 + 4 \cdot (-2) + 1 \cdot 1 \\ 1 \cdot 1 + 4 \cdot 2 + 2 \cdot (-2) + (-2) \cdot 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \\ 3 \end{bmatrix}$$

☐

6.

Proof. We choose $u_1 = u_2 = u_3 = 1$ and $u_4 = 2$. Then $x_3 = -5x_4 - u_3 - u_2 + 3u_1 = -5x_4 + 1$, $x_2 = 7x_4 + u_4 - 3u_1 = 7x_4 - 1$, and $x_1 = u_1 - x_2 - 2x_3 - 3x_4 = 1 - (7x_4 - 1) - 2(-5x_4 + 1) - 3x_4 = 0$. ☐

7.

Proof.

$$\begin{aligned} & [1, -2, -1, 1] \begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 1 \\ 2 & 1 & 2 & 3 \\ 3 & 4 & 6 & 2 \end{bmatrix} \\ = \; & [1 \cdot 1 - 2 \cdot 1 - 1 \cdot 2 + 1 \cdot 3, 1 \cdot 1 - 2 \cdot 2 - 1 \cdot 1 + 1 \cdot 4, 1 \cdot 2 - 2 \cdot 3 - 1 \cdot 2 + 1 \cdot 6, 1 \cdot 3 - 2 \cdot 1 - 1 \cdot 3 + 1 \cdot 2] \\ = \; & 0. \end{aligned}$$

☐

8.

Proof. Suppose a row vector $x = (x_1, x_2, x_3, x_4)$ satisfies $xM = 0$. Then we can proceed according to Gaussian elimination

$$\begin{cases} x_1 + x_2 + 2x_3 + 3x_4 = 0 \\ x_2 + 2x_2 + x_3 + 4x_4 = 0 \\ 2x_1 + 3x_2 + 2x_3 + 6x_4 = 0 \\ 3x_1 + x_2 + 3x_3 + 2x_4 = 0 \end{cases} \Rightarrow \begin{cases} x_2 - x_3 + x_4 = 0 \\ x_2 - 2x_3 = 0 \\ -2x_2 - 3x_3 - 7x_4 = 0 \end{cases} \Rightarrow \begin{cases} -x_3 - x_4 = 0 \\ -5x_3 - 5x_4 = 0. \end{cases}$$

So we have $x_1 = x_4$, $x_2 = -2x_4$, and $x_3 = -x_4$, i.e. $x = x_4(1, -2, -1, 1)$, a multiple of l in Exercise 7. Equation (22) has a solution if and only if $u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$ is in R_M. By Theorem 5 of Chapter 3, this is equivalent to $yu = 0$, $\forall y \in N_{M'}$ (elements of $N_{M'}$ are seen as row vectors). We have proved y is a multiple of l. Hence condition (23), which is just $lu = 0$, is sufficient for the solvability of the system (22). ☐

5 Determinant and Trace

1. (i)

Proof. By formula (5), we have $|P(x_1, \cdots, x_n)| = |\prod_{i<j}(x_i - x_j)| = |\prod_{i\neq j}(x_i - x_j)| = |P(p(x_1, \cdots, x_n))|$. By formula (6), we have $|P(p(x_1, \cdots, x_n)| = |\sigma(p)||P(x_1, \cdots, x_n)|$. Combined, we conclude $|\sigma(p)| = 1$. □

(ii)

Proof. By definition, we have

$$P(p_1 \circ p_2(x_1, \cdots, x_n)) = P(p_1(p_2(x_1, \cdots, x_n))) = \sigma(p_1)P(p_2(x_1, \cdots, x_2)) = \sigma(p_1)\sigma(p_2)P(x_1, \cdots, x_n).$$

So $\sigma(p_1 \circ p_2) = \sigma(p_1)\sigma(p_2)$. □

2.

Proof. To see (c) is true, we suppose t interchange i_0 and j_0. Without loss of generality, we assume $i_0 < j_0$. Then

$$\begin{aligned} P(t(x_1, \cdots, x_n)) &= P(x_1, \cdots, x_{j_0}, \cdots, x_{i_0}, \cdots, x_n) \\ &= (x_{j_0} - x_{i_0}) \prod_{i<j, (i,j)\neq(i_0,j_0)} (x_i - x_j) \\ &= -\prod_{i<j}(x_i - x_j) \\ &= -P(x_1, \cdots, x_n). \end{aligned}$$

So $\sigma(t) = -1$.

To see (d) is true, note formula (9) is equivalent to $\mathrm{id} = t_k \circ \cdots t_1 \circ p^{-1}$. Acting these operations on $(1, \cdots, n)$, we have $(1, \cdots, n) = t_k \circ \cdots \circ t_1(p^{-1}(1), \cdots, p^{-1}(n))$. Then the problem is reduced to proving that a sequence of transpositions can sort an array of numbers into ascending order. There are many ways to achieve that. For example, we can let t_1 be the transposition that interchanges $p^{-1}(1)$ and $p^{-1}(i_0)$, where i_0 satisfies $p^{-1}(i_0) = 1$. That is, t_1 puts 1 in the first position of the sequence. Then we let t_2 be the transposition that puts 2 to the second position. We continue this procedure until we sort out the whole sequence. This shows sorting can be accomplished by a sequence of transpositions. □

3.

Proof. For any transposition t, we have $t \circ t = \mathrm{id}$. So if $p = t_k \circ \cdots \circ t_1$, we can get another decomposition $p = t_k \circ \cdots \circ t_1 \circ t \circ t$. This shows the decomposition is not unique.

Suppose the permutation p has two different decompositions into transpositions: $p = t_k \circ \cdots \circ t_1 = t'_m \circ \cdots \circ t'_1$. By formula (7), part (b) and formula (8), $\sigma(p) = (-1)^k = (-1)^m$. So $k - m$ is an even number. This shows the parity of the member of factors is unique. □

4.

Proof. To verify Property (ii), note for any index j, $\alpha, \beta \in K$, we have

$$\begin{aligned} & D(a_1, \cdots, \alpha a'_j + \beta a_j, \cdots, a_n) \\ = & \sum \sigma(p) a_{p_1 1} \cdots (\alpha a'_{p_j j} + \beta a'_{p_j j}) \cdots a_{p_n n} \\ = & \sum [\alpha \sigma(p) a_{p_1 1} \cdots a'_{p_j j} \cdots a_{p_n n} + \beta \sigma(p) a_{p_1 1} \cdots a_{p_j j} \cdots a_{p_n n}] \\ = & \alpha D(a_1, \cdots, a'_j, \cdots, a_n) + \beta D(a_1, \cdots, a_j, \cdots, a_n). \end{aligned}$$

To verify Property (iii), note $e_{p_1 1} \cdots e_{p_n n}$ is non-zero if and only if $p_i = i$ for any $1 \leq i \leq n$. In this case the product is 1.

To verify Property (iv), note for any $i \neq j$, if we denote by t the transposition that interchanges i and j, then $p \mapsto p \circ t$ is a one-to-one and onto map from the set of all permutations to itself. Therefore, we have

$$\begin{aligned}
& D(a_1, \cdots, a_i, \cdots, a_j, \cdots, a_n) \\
= & \sum \sigma(p) a_{p_1 1} \cdots a_{p_i i} \cdots a_{p_j j} \cdots a_{p_n n} \\
= & \sum (-1)\sigma(p \circ t) a_{p\circ t_1 1} \cdots a_{p\circ t_j i} \cdots a_{p\circ t_i j} \cdots a_{p\circ t_n n} \\
= & \sum (-1)\sigma(p \circ t) a_{p\circ t_1 1} \cdots a_{p\circ t_i j} \cdots a_{p\circ t_j i} \cdots a_{p\circ t_n n} \\
= & (-1) \sum \sigma(q) a_{q_1 1} \cdots a_{q_i j} \cdots a_{q_j i} \cdots a_{q_n n} \\
= & -D(a_1, \cdots, a_j, \cdots, a_i, \cdots, a_n).
\end{aligned}$$

□

5.

Proof. By property (iv), $D(a_1, \cdots, a_i, \cdots, a_i, \cdots, a_n) = -D(a_1, \cdots, a_i, \cdots, a_i, \cdots, a_n)$. So add to both sides of the equations $D(a_1, \cdots, a_i, \cdots, a_i, \cdots, a_n)$, we have $2D(a_1, \cdots, a_i, \cdots, a_i, \cdots, a_n) = 0$. If the character of the field K is not two, we can conclude $D(a_1, \cdots, a_i, \cdots, a_i, \cdots, a_n) = 0$. □

Remark 5. *This exercise and Exercise 5.4 together show formula (16) is equivalent to Properties (i)-(iii), provided the character of K is not two. Therefore, for $K = \mathbb{R}$ or $\mathbb{C}$, we can either use (16) or properties (i)-(iii) as the definition of determinant.*

6.

Proof. If two column vectors a_i and a_j $(i \neq j)$ of A_{11} are equal, we have $\begin{bmatrix} 0 \\ a_i \end{bmatrix} = \begin{bmatrix} 0 \\ a_j \end{bmatrix}$. So $C(A_{11}) = 0$ and property (i) is satisfied. Since any linear operation on a column vector a_i of A_{11} can be naturally extended to $\begin{bmatrix} 0 \\ a_i \end{bmatrix}$, property (ii) is also satisfied. Finally, we note when $A_{11} = I_{(n-1)\times(n-1)}$, $\begin{bmatrix} 1 & 0 \\ 0 & A_{11} \end{bmatrix} = I_{n\times n}$. So property (iii) is satisfied. □

7.

Proof. We first move the j-th column to the position of the first column. This can be done by interchanging neighboring columns $(j-1)$ times. The determinant of the resulted matrix A_1 is $(-1)^{j-1}\det A$. Then we move the i-th row to the position of the first row. This can be done by interchanging neighboring rows $(i-1)$ times. The resulted matrix A_2 has a determinant equal to $(-1)^{i-1}\det A_1 = (-1)^{i+j}\det A$. On the other hand, A_2 has the form of $\begin{pmatrix} 1 & * \\ 0 & A_{ij} \end{pmatrix}$. By Lemma 4, we have $\det A_{ij} = \det A_2 = (-1)^{i+j}\det A$. So $\det A = (-1)^{i+j}\det A_{ij}$. □

Remark 6. *Rigorously speaking, we only proved that swapping two neighboring columns will give a minus sign to the determinant (Property (iv)), but we haven't proved this property for neighboring rows. This can be made rigorous by using $\det A = \det A^T$ (Exercise 8 of this chapter).*

8.

Proof. We first show for any permutation p, $\sigma(p) = \sigma(p^{-1})$. Indeed, by formula (7)(b), we have $1 = \sigma(id) = \sigma(p \circ p^{-1}) = \sigma(p)\sigma(p^{-1})$. By formula (7)(a), we conclude $\sigma(p) = \sigma(p^{-1})$. Second, we denote by b_{ij} the

(i, j)-th entry of A^T. Then $b_{ij} = a_{ji}$. By formula (16) and the fact that $p \mapsto p^{-1}$ is a one-to-one and onto map from the set of all permutations to itself, we have

$$\begin{aligned}
\det A^T &= \sum \sigma(p) b_{p_1 1} \cdots b_{p_n n} \\
&= \sum \sigma(p) a_{1 p_1} \cdots a_{n p_n} \\
&= \sum \sigma(p^{-1}) a_{(p^{-1}\circ p)_1 p_1} \cdots a_{(p^{-1}\circ p)_n p_n} \\
&= \sum \sigma(p^{-1}) a_{p^{-1}(p_1) p_1} \cdots a_{p^{-1}(p_n) p_n} \\
&= \sum \sigma(p^{-1}) a_{p^{-1}(1) 1} \cdots a_{p^{-1}(n) n} \\
&= \det A.
\end{aligned}$$

□

9.

Proof. By Exercise 2, it suffices to prove the property for transpositions. Suppose p interchanges i_1, i_2 and q interchanges j_1, j_2. Denote by P and Q the corresponding permutation matrices, respectively. Then for any $x = (x_1, \cdots, x_n)^T \in \mathbb{R}^n$, we have ($\delta_{ij}$ is the Kronecker sign)

$$(Px)_i = \sum P_{ij} x_j = \sum \delta_{p(i) j} x_j = \begin{cases} x_{i_2} & \text{if } i = i_1 \\ x_{i_1} & \text{if } i = i_2 \\ x_i & \text{otherwise.} \end{cases}$$

This shows the action of P on any column vector x performs the permutation p on the components of x. Similarly, we have

$$(Qx)_i = \begin{cases} x_{j_2} & \text{if } i = j_1 \\ x_{j_1} & \text{if } i = j_2 \\ x_i & \text{otherwise.} \end{cases}$$

Since $(PQ)(x) = P(Q(x))$, the action of matrix PQ on x performs first the permutation q and then the permutation p on the components of x. Therefore, the permutation matrix associated with $p \circ q$ is the product of P and Q. □

10.

Proof.

$$\operatorname{tr}(AB) = \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{m} \sum_{i=1}^{n} a_{ji} b_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} a_{ji} = \sum_{i=1}^{n} (BA)_{ii} = \operatorname{tr}(BA),$$

where the third equality is obtained by interchanging the names of the indices i, j. □

11.

Proof.

$$\operatorname{tr}(AA^T) = \sum_i (AA^T)_{ii} = \sum_i \sum_j A_{ij} A^T_{ji} = \sum_i \sum_j a_{ij} a_{ij} = \sum_{ij} a_{ij}^2.$$

□

12.

Proof. Apply Laplace expansion of a determinant according to its columns (Theorem 6). □

13.

Proof. Apply Laplace expansion of a determinant according to its columns (Theorem 6) and work by induction. □

14.

Proof. Denote by $M(n)$ the number of multiplications needed to evaluate $\det A$ of an $n \times n$ matrix A by using Gaussian elimination to bring it into upper triangular form. To use the first row to eliminate a_{21}, a_{31}, $\cdots$, a_{n1}, we need $n(n-1)$ multiplications. So $M(n) = n(n-1) + M(n-1)$ with $M(1) = 0$. So $M(n) = \sum_{k=1}^{n} k(k-1) = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} = \frac{(n-1)n(n+1)}{3}$. □

15.

Proof. Denote by $M(n)$ the number of multiplications needed to evaluate the determinant of an $n \times n$ matrix by formula (16). Then $M(n) = nM(n-1)$. So $M(n) = n!$. □

16.

Proof. We apply Laplace expansion of a determinant according to its columns (Theorem 6):

$$\begin{aligned}
\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} &= a \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - d \det \begin{bmatrix} b & c \\ h & i \end{bmatrix} + g \det \begin{bmatrix} b & c \\ e & f \end{bmatrix} \\
&= a(ie - fh) - d(ib - ch) + g(bf - ce) \\
&= aei + bfg + cdh - gec - afh - idb.
\end{aligned}$$

□

6 Spectral Theory

1.

Proof. $f_{32} = a_1^{32}/\sqrt{5} = 2178309$. □

2. (a)

Proof. Denote by h_j the eigenvector corresponding to the eigenvalue a_j. For any $h \in \mathbb{C}^n$, there exist $\theta_1, \cdots, \theta_n \in \mathbb{C}$ such that $h = \sum_j \theta_j h_j$. So $A^N h = \sum_j \theta_j a_j^N h_j$. Define $b = \max\{|a_1|, \cdots, |a_n|\}$. Then for any $1 \le k \le n$, $|(A^N h)_k| = |\sum_j \theta_j a_j^N (h_j)_k| \le b^N \sum_j |\theta_j||(h_j)_k| \to 0$, as $N \to \infty$, since $0 \le b < 1$. This shows $A^N h \to 0$ as $N \to \infty$. □

(b)

Proof. We use the same notation as in part (a). Since $h \neq 0$, there exists some k_0 so that $h_{k_0} = \sum_j \theta_j (h_j)_{k_0} \neq 0$. Then $|(A^N h)_{k_0}| = |\sum_j \theta_j a_j^N (h_j)_{k_0}|$. Define $b_1 = \max_{1 \le i \le n}\{|a_i| : \theta_i \neq 0, (h_i)_{k_0} \neq 0\}$. Then $b_1 > 1$ and hence $|(A^N h)_{k_0}| = |b_1|^N \left| \sum_{i=1}^{n} \theta_i \frac{a_i^N}{b_1^N} (h_i)_{k_0} \right| \to \infty$ as $N \to \infty$. □

3.

Proof. The verification is straightforward. □

4.

Proof. Formula (24) gives us $Af = af + h$, which is formula (25) when $N = 1$. Suppose (25) holds for any $n \le N$, then $A^{N+1} f = A(A^N f) = A(a^N f + Na^{N-1} h) = a^N Af + Na^{N-1} Ah = a^N(af + h) + Na^{N-1} ah = a^{N+1} f + (N+1)a^N h$. So (25) also holds for $N+1$. By induction, (25) holds for any $N \in \mathbb{N}$. □

5.

Proof. Suppose $q(s) = \sum_{i=0}^{n} b_i s^i$, then by formula (25), $q(A)f = \sum_{i=0}^{n} b_i A^i f = \sum_{i=0}^{n} b_i (a^i f + i a^{i-1} h) = (\sum_{i=0}^{n} b_i a^i) f + (\sum_{i=1}^{n} i b_i a^{i-1}) h = q(a) f + q'(a) h$. $\square$

6.

Proof. By Lemma 9, $N_{p_1 \cdots p_k} = N_{p_1} \oplus N_{p_2 \cdots p_k} = N_{p_1} \oplus (N_{p_2} \oplus N_{p_3 \cdots p_k}) = N_{p_1} \oplus N_{p_2} \oplus N_{p_3 \cdots p_k} = \cdots = N_{p_1} \oplus N_{p_2} \oplus \cdots \oplus N_{p_k}$. $\square$

7.

Proof. For any $x \in N_d(a)$, we have $(A - aI)^d (Ax) = (A - aI)^{d+1} x + a(A - aI)^d x = 0$. So $Ax \in N_d(a)$. $\square$

8.

Proof. A number is an eigenvalue of A if and only if it's a root of the characteristic polynomial p_A. So $p_A(s)$ can necessarily be written as $p_A(s) = \prod_1^k (s - a_i)^{m_i}$ with each m_i a positive integer ($i = 1, \cdots, k$). We have shown in the text that p_A is a multiple of m_A, so we can assume $m_A(s) = \prod_{i=1}^k (s - a_i)^{r_i}$ with each r_i satisfying $0 \le r_i \le m_i$ ($i = 1, \cdots, k$). We argue $r_i = d_i$ for any $1 \le i \le k$.

Indeed, if for some j, $r_j < d_j$, we can find $x \in N_{d_j}(a_j) \setminus N_{r_j}(a_j)$ with $x \neq 0$. Define $q(s) = \prod_{i=1, i \neq j}^k (s - a_i)^{r_i}$, then by Corollary 10 x can be uniquely decomposed into $x' + x''$ with $x' \in N_q$ and $x'' \in N_{r_j}(a_j)$. We have $0 = (A - a_j I)^{d_j} x = (A - a_j I)^{d_j} x' + 0$. So $x' \in N_q \cap N_{d_j}(a_j) = \{0\}$. This implies $x = x'' \in N_{r_j}(a_j)$. Contradiction. Therefore, $r_i \ge d_i$ for any $1 \le i \le k$.

On the other hand, if for some j, $r_j > d_j$, then $m'_A = \prod_{i=1, i \neq j}^k (s - a_i)^{r_i} \cdot (s - a_j)^{d_j}$ satisfies $\deg(m'_A) < \deg(m_A)$ and $N_{m'_A} = N_{r_1}(a_1) \oplus \cdots \oplus N_{d_j}(a_j) \oplus \cdots \oplus N_{r_k}(a_k) = N_{d_1}(a_1) \oplus \cdots \oplus N_{d_k}(a_k) = \mathbb{C}^n$ by $r_i \ge d_i$ ($1 \le i \le k$), which is proved above. So m_A cannot be the minimal polynomial. Contradiction.

Combined, we conclude $m_A(s) = \prod_{i=1}^k (s - a_i)^{d_i}$. $\square$

Remark 7. *Minimal polynomial is defined from the algebraic point of view as the generator of* $\{$*polynomial* $p : p(A) = 0\}$. *So the powers of its linear factors are given algebraically. Meanwhile, the index of an eigenvalue is defined from the geometric point of view. Theorem 11 says they are equal.*

Remark 8. *An alternative proof goes as follows. By Theorem 5, Corollary 10, and the definition of index, we have (we assume the characteristic polynomial* $p_A(s) = \prod_{j=1}^k (s - a_j)^{m_j}$*)*

$$\mathbb{C}^n = N_{p_A} = \bigoplus_{j=1}^k N_{m_j}(a_j).$$

$(A - a_j I)$ *is nilpotent on* $N_{m_j}(a_j)$ *with index* d_j. *So* $A|_{N_{m_j}(a_j)}$ *has minimal polynomial* $(s - a_j)^{d_j}$ *and the minimal polynomial of* A *on* $\mathbb{C}^n$ *is therefore* $\prod_1^k (s - a_j)^{d_j}$.

Remark 9. *As a corollary, we claim* A *can be diagonalized over the field* $\mathbb{F}$ *if and only if its minimal polynomial can be decomposed into the product of distinct polynomials of degree 1 over the field* $\mathbb{F}$. *Indeed, by the uniqueness of minimal polynomial, we have*

$$\begin{aligned} & m_A \text{ is the product of distinct polynomials of degree 1} \\ \iff & d_1 = \cdots = d_k = 1 \\ \iff & \mathbb{C}^n = \bigoplus_{j=1}^k N_1(a_j) \\ \iff & A \text{ can be diagonalized, where the basis is } \cup_{j=1}^k \{e_{j1}, \cdots, e_{ji_j}\} \text{ with } \{e_{j1}, \cdots, e_{ji_j}\} \text{ a basis for } N_1(a_j) \end{aligned}$$

Remark 10. *The following proposition gives an elementary proof to Lemma 10, Chapter 9.*

Proposition 1 (Geometric multiplicity and algebraic multiplicity). *Let A be an $n \times n$ matrix over a field $\mathbb{F}$ and a an eigenvalue of A. If m is the multiplicity of a as a root of the characteristic polynomial of A, then $dim N_1(a) \leq m$, where $N_1(a) =$ null space of $(A - aI)$ is the space of eigenvectors pertaining to a.*

m is called the algebraic multiplicity of a and $dim N_1(a)$ is called the geometric multiplicity of a. So this theorem says "geometric multiplicity $\leq$ algebraic multiplicity".

Proof. Let $v_1, \cdots, v_s$ be a basis of $N_1(a)$ and extend it to a basis of $\mathbb{F}^n$: $v_1, \cdots, v_s, u_1, \cdots, u_r$. Define $U = (v_1, \cdots, v_s, u_1, \cdots, u_r)$. Then

$$\begin{aligned} U^{-1}AU &= U^{-1}A(v_1, \cdots, v_s, u_1, \cdots, u_r) \\ &= U^{-1}(av_1, \cdots, av_s, Au_1, \cdots, Au_r) \\ &= (aU^{-1}v_1, \cdots, aU^{-1}v_s, U^{-1}Au_1, \cdots, U^{-1}Au_r). \end{aligned}$$

Because $U^{-1}U = I$, $U^{-1}AU = \begin{bmatrix} aI_{s\times s} & B \\ 0 & C \end{bmatrix}$ and $\det(\lambda I - A) = \det(\lambda I - U^{-1}AU) = (\lambda - a)^s \det(\lambda I - C)$.[1] So $s \leq m$. □

Remark 11. *It's a good place to summarize the relationship between index, algebraic multiplicity, geometric multiplicity, and the dimension of the space of generalized eigenvectors pertaining to a given eigenvalue. We have the following sequence of results:*

Proposition 2 (Index and algebraic multiplicity). *Let A be an $n \times n$ matrix over a field $\mathbb{F}$ and a an eigenvalue of A. Denote by m the multiplicity of a as a root of the characteristic polynomial p_A and by d the index of a, then $d \leq m$.*

Proof. Let $q(s) = p_A(s)/(s-a)^m$. Then $\mathbb{C}^n = N_{p_A} = N_q \bigoplus N_m(a)$. For any $v \in N_{m+1}(a)$, v can be uniquely written as $v = v' + v''$ with $v' \in N_q$ and $v'' \in N_m(a)$. Then $v' = v - v'' \in N_{m+1}(a) \cap N_q$. Similar to the second part of the proof of Lemma 9, we can show $v' = 0$. So $v = v'' \in N_m(a)$. This shows $N_{m+1}(a) = N_m(a)$ and hence $d \leq m$. □

Proposition 3 (Algebraic multiplicity and the dimension of the space of generalized eigenvectors). *Let A be an $n \times n$ matrix over a field $\mathbb{F}$ and a an eigenvalue of A. Denote by m the multiplicity of a as a root of the characteristic polynomial p_A and by k the dimension of the space of generalized eigenvectors of A pertaining to a, then $k = m$.*

Proof. See Theorem 11 of Chapter 9. □

Summary: numbers associated with an eigenvalue. *Let A be an $n \times n$ matrix over a field $\mathbb{F}$ and a an eigenvalue of A. Then we have*

$$\begin{aligned} \begin{cases} \text{geometric multiplicity of } a \\ \text{index of } a \end{cases} &\leq \text{dim. of the space of generalized eigenvectors pertaining to } a \\ &= \text{algebraic multiplicity of } a. \end{aligned}$$

9.

Proof. The extension is straightforward as the key feature of the proof, "B maps $N^{(j)}$ into $N^{(j)}$", remains the same regardless of the number of linear maps, as far as they commute pairwise. □

10.

Proof. For any $i \in \{1, \cdots, n\}$, by Theorem 17, $(l_i, x_j) = 0$ for any $j \neq i$. Since $x_1, \cdots, x_n$ span the whole space and $l_i \neq 0$, we must have$(l_i, x_i) \neq 0$, $i = 1, \cdots, n$. This proves (a) of Theorem 18. For (b), we note if $x = \sum_{j=1}^k k_j x_j$, then $(l_i, x) = k_i(l_i, x_i)$. So $k_i = (l_i, x)/(l_i, x_i)$. □

[1] For the last equality, see, for example, Munkres [5], page 24, Problem 6.

11. (a)

Proof. The matrix is symmetric, so it's equal to its transpose and the eigenvectors are the same: for eigenvalue $a_1 = \frac{1+\sqrt{5}}{2}$, the eigenvector is $h_1 = \begin{bmatrix} 1 \\ a_1 \end{bmatrix}$; for eigenvalue $a_2 = \frac{1-\sqrt{5}}{2}$, the eigenvector is $h_2 = \begin{bmatrix} 1 \\ a_2 \end{bmatrix}$. □

(b)

Proof. We note $(h_1, h_1) = 1 + a_1^2 = \frac{5+\sqrt{5}}{2}$ and $(h_2, h_2) = 1 + a_2^2 = \frac{5-\sqrt{5}}{2}$. For $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we have $(h_1, x) = a_1$ and $(h_2, x) = a_2$. So using formula (44) and (45), $x = c_1 h_1 + c_2 h_2$ with

$$c_1 = a_1 / \frac{5+\sqrt{5}}{2} = 1/\sqrt{5},\ c_2 = a_2 / \frac{5-\sqrt{5}}{2} = -1/\sqrt{5}.$$

This agrees with the expansion obtained in Example 2. □

12.

Proof. The transpose of the matrix has the same eigenvalues $a_1 = 2$, $a_2 = 5$. Solving the equation $\begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix}$, we have $l_1 = \begin{bmatrix} 1 & -1 \end{bmatrix}$. Solving the equation $\begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5 \begin{bmatrix} x \\ y \end{bmatrix}$, we have $l_2 = \begin{bmatrix} 1 & 2 \end{bmatrix}$. Then it's easy to calculate $(l_1, h_1) = 3$, $(l_1, h_2) = 0$, $(l_2, h_1) = 0$, and $(l_2, h_2) = 3$. □

13.

Proof.

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda & -1 & -1 \\ -1 & \lambda & -1 \\ -1 & -1 & \lambda \end{bmatrix} = \det \begin{bmatrix} 0 & -1-\lambda & -1+\lambda^2 \\ 0 & \lambda+1 & -1-\lambda \\ -1 & -1 & \lambda \end{bmatrix} = -[(\lambda+1)^2 - (\lambda^2-1)(\lambda+1)] = (\lambda+1)^2(\lambda-2).$$

So the eigenvalues of A are -1 and 2, and the eigenvalue 2 has a multiplicity of 2. □

7 Euclidean Structure

1.

Proof. By letting $y = \frac{x}{\|x\|}$, we get $\|x\| \le \max_{\|y\|=1}(x, y)$. By Schwartz Inequality, $\max_{\|y\|=1}(x, y) \le \|x\|$. Combined, we must have $\|x\| = \max_{\|y\|=1}(x, y)$. □

2.

Proof. $\forall x, y$ and suppose their decomposition are $x_1 + x_2$, $y_1 + y_2$, respectively. Here $x_1, y_1 \in Y$ and $x_2, y_2 \in Y^\perp$. Then $(P_Y^* y, x) = (y, P_Y x) = (y_1 + y_2, x_1) = (y_1, x_1) = (y_1, x) = (P_Y y, x)$. By the arbitrariness of x and y, $P_Y = P_Y^*$. □

3.

Proof. Under the reflection across the plane $\{(x_1, x_2, x_3) : x_3 = 0\}$, point (x_1, x_2, x_3) will be mapped to $(x_1, x_2, -x_3)$. So the corresponding matrix is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, whose determinant is -1. □

4.

Proof. Suppose the plane L is determined by the equation $Ax + By + Cz = D$. For any point $x = (x_1, x_2, x_3)' \in \mathbb{R}^3$, we first find $y = (y_1, y_2, y_3)' \in L$ such that the line segment $xy \perp L$. Then y must satisfy the following equations

$$\begin{cases} Ay_1 + By_2 + Cy_3 = D \\ (y_1 - x_1, y_2 - x_2, y_3 - x_3) = k(A, B, C) \end{cases}$$

where k is some constant. Solving the equations gives us $k = \frac{D-(Ax_1+Bx_2+Cx_3)}{A^2+B^2+C^2}$ and

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + k \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \frac{1}{A^2+B^2+C^2} \begin{bmatrix} A^2 & AB & AC \\ AB & B^2 & BC \\ CA & CB & C^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \frac{D}{A^2+B^2+C^2} \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

So the symmetric point $z = (z_1, z_2, z_3)'$ of x with respect to L is given by

$$\begin{aligned} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} &= 2\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \frac{2}{A^2+B^2+C^2} \begin{bmatrix} A^2 & AB & AC \\ AB & B^2 & BC \\ CA & CB & C^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \frac{2D}{A^2+B^2+C^2} \begin{bmatrix} A \\ B \\ C \end{bmatrix} \\ &= \frac{1}{A^2+B^2+C^2} \begin{bmatrix} -A^2+B^2+C^2 & -2AB & -2AC \\ -2AB & A^2-B^2+C^2 & -2BC \\ -2CA & -2CB & A^2+B^2-C^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \frac{2D}{A^2+B^2+C^2} \begin{bmatrix} A \\ B \\ C \end{bmatrix}. \end{aligned}$$

To make the reflection R a linear mapping, it's necessary and sufficient that $D = 0$. So the problem's statement should be corrected to "let R be reflection across any plane in $\mathbb{R}^3$ that contains the origin". Then

$$R = \frac{1}{A^2+B^2+C^2} \begin{bmatrix} -A^2+B^2+C^2 & -2AB & -2AC \\ -2AB & A^2-B^2+C^2 & -2BC \\ -2CA & -2CB & A^2+B^2-C^2 \end{bmatrix}.$$

R is symmetric, so $R^* = R$ and by plain calculation, we have $R^*R = R^2 = I$. By Theorem 12, R is an isometry. □

5.

Proof. Suppose M is an $n \times n$ orthogonal matrix. Let $r_1, \cdots, r_n$ be its column vectors. Then

$$I = MM^T = \begin{bmatrix} r_1 \\ \cdots \\ r_n \end{bmatrix} [r_1^T, \cdots, r_n^T] = \begin{bmatrix} r_1 r_1^T & r_1 r_2^T & \cdots & r_1 r_n^T \\ \cdots & \cdots & \cdots & \cdots \\ r_n r_1^T & r_n r_2^T & \cdots & r_n r_n^T \end{bmatrix}.$$

So M is orthogonal if and only if $r_i r_j^T = \delta_{ij}$ $(1 \le i, j \le n)$. □

6.

Proof. Note $|a_{ij}| = \text{sign}(a_{ij}) \cdot e_i^T A e_j$, where e_k is the column vector that has 1 as the k-th entry and 0 elsewhere. Then we apply (ii) of Theorem 13. □

7.

Proof. See the solution in the textbook. □

8.

Proof. For any $x, y \in X$ and $a \in \mathbb{C}$, $0 \le \|x - ay\| = \|x\|^2 - 2\text{Re}(x, ay) + |a|^2\|y\|^2$. Let $a = \frac{(x,y)}{\|y\|^2}$ (assume $y \ne 0$), then we have

$$0 \le \|x\|^2 - 2\text{Re}\left\{ \frac{\overline{(x,y)}}{\|y\|^2}(x,y) \right\} + \frac{|(x,y)|^2}{\|y\|^2},$$

which gives after simplification $|(x,y)| \le \|x\|\|y\|$. □

9.

Proof. Proofs are the same as the ones for the real Euclidean space. ☐

10.

Proof. Proof is the same as the one for real Euclidean space. ☐

11.

Proof. If M is a unitary map, then by parallelogram law, M preserves inner product. So $\forall x, y$, $(x, M^*My) = (Mx, My) = (x, y)$. Since x is arbitrary, $M^*My = y$, $\forall y \in X$. So $M^*M = I$. Conversely, if $M^*M = I$, $(x, x) = (x, M^*Mx) = (Mx, Mx)$. So M is an isometry. ☐

12.

Proof. $(M^{-1}x, M^{-1}x) = (M(M^{-1}x), M(M^{-1}x)) = (x, x)$. $(Mx, Mx) = (x, x) = (M^*Mx, M^*Mx)$. By $R_M = X$, $(y, y) = (M^*y, M^*y)$, $\forall y \in X$. So M^{-1} and M^* are both unitary. ☐

13.

Proof. If M, N are two unitary maps, then $(MN)^*(MN) = N^*M^*MN = N^*N = I$. So the set of unitary maps is closed under multiplication. Exercise 12 shows that each unitary map has a unitary inverse. So the set of unitary maps is a group under multiplication. ☐

14.

Proof. By Exercise 8 of Chapter 5, $\det M^* = \det \overline{M}^T = \overline{\det M}$. So by $M^*M = I$, we have $1 = \det M^* \det M = |\det M|^2$, i.e. $|\det M| = 1$. ☐

15.

Proof. $(Mf, Mf) = \int_{-1}^{1} Mf(s)\overline{Mf(s)}ds = \int_{-1}^{1} m(s)f(s)\bar{m}(s)\bar{f}(s)ds = \int_{-1}^{1} |m(s)|^2|f(s)|^2 ds = (f, f)$. This shows M is unitary. ☐

16.

Proof. The proof is very similar to that of real case, so we omit the details. Note we need the complex version of Schwartz inequality (Exercise 8). ☐

17.

Proof. We have

$$(AA^*)_{ij} = [a_{i1}, \cdots, a_{in}] \begin{bmatrix} \bar{a}_{j1} \\ \cdots \\ \bar{a}_{jn} \end{bmatrix} = \sum_{k=1}^{n} a_{ik}\bar{a}_{jk}.$$

So $(AA^*)_{ii} = \sum_{k=1}^{n} |a_{ik}|^2$ and $\operatorname{tr}(AA^*) = \sum_{i,j} |a_{ij}|^2$. ☐

18.

Proof. This is straightforward from the result of Exercise 17. ☐

19.

Proof. Suppose λ_1 and λ_2 are two eigenvalues of A. Then by Theorem 3 of Chapter 6, $\lambda_1 + \lambda_2 = \operatorname{tr}A = 4$ and $\lambda_1\lambda_2 = \det A = 3$. Solving the equations gives us $\lambda_1 = 1$, $\lambda_2 = 3$. By formula (46), $\|A\| \geq 3$. According to formula (51), we have $\|A\| \leq \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$. Combined, we have $3 \leq \|A\| \leq \sqrt{14} \approx 3.7417$. ☐

20. (i)

Proof. For any α_1, $\alpha_2 \in \mathbb{F}$, we have

$$\begin{aligned}(w(\alpha_1 x_1 + \alpha_2 x_2, y), z) &= \det(\alpha_1 x_1 + \alpha_2 x_2, y, z) = \alpha_1 \det(x_1, y, z) + \alpha_2 \det(x_2, y, z) \\ &= \alpha_1 (w(x_1, y), z) + \alpha_2 (w(x_2, y), z) = (\alpha_1 w(x_1, y) + \alpha_2 w(x_2, y), z).\end{aligned}$$

Since z is arbitrary, we necessarily have $w(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 w(x_1, y) + \alpha_2 w(x_2, y)$. Similarly, we can prove $w(x, \alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 w(x, y_1) + \alpha_2 w(x, y_2)$. Combined, we have proved w is a bilinear function of x and y. □

(ii)

Proof. We note

$$(w(x, y), z) = \det(x, y, z) = -\det(y, x, z) = -(w(y, x), z) = (-w(y, x), z).$$

By the arbitrariness of z, we conclude $w(x, y) = -w(y, x)$, i.e. $y \times x = -x \times y$. □

(iii)

Proof. Since $(w(x, y), x) = \det(x, y, x) = 0$ and $(w(x, y), y) = \det(x, y, y) = 0$, $x \times y$ is orthogonal to both x and y. □

(iv)

Proof. We suppose every vector is in column form and R is the matrix that represents a rotation. Then

$$(Rx \times Ry, z) = \det(Rx, Ry, z) = (\det R) \cdot \det(x, y, R^{-1} z)$$

and

$$(R(x \times y), z) = (R(x \times y))^T z = (x \times y)^T R^T z = (x \times y, R^T z) = \det(x, y, R^T z).$$

A rotation is isometric, so $R^T = R^{-1}$ and $\det R = \pm 1$. Combing the above two equations gives us $\pm(Rx \times Ry, z) = (R(x \times y), z)$. Since z is arbitrary, we must have $\pm Rx \times Ry = R(x \times y)$. □

(v)

Proof. In the equation $\det(x, y, z) = (x \times y, z)$, we set $z = x \times y$. Since the geometrical meaning of $\det(x, y, z)$ is the signed volume of a parallelogram determined by x, y, z, and since $z = x \times y$ is perpendicular to x and y, we have $\det(x, y, z) = \pm\|x\|\|y\| \sin\theta \|z\|$, where θ is the angle between x and y. Then by $(x \times y, z) = \|z\|^2$, we conclude $\|x \times y\| = \|z\| = \pm\|x\|\|y\| \sin\theta$. □

(vi)

Proof.

$$1 = \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = (\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}).$$

So $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ 1 \end{bmatrix}$. By part (iii), we necessarily have $a = b = 0$. Therefore, we can conclude $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. □

(vii)

Proof. By Exercise 16 of Chapter 5,

$$\begin{aligned}
\det\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} &= \det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \\
&= aei + bfg + cdh - gec - hfa - idb \\
&= (bf - ec)g + (cd - fa)h + (ae - db)i \\
&= \begin{bmatrix} bf - ce & cd - af & ae - bd \end{bmatrix} \begin{bmatrix} g \\ h \\ i \end{bmatrix}.
\end{aligned}$$

So we have

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \times \begin{bmatrix} d \\ e \\ f \end{bmatrix} = \begin{bmatrix} bf - ce \\ cd - af \\ ae - bd \end{bmatrix}.$$

□

21.

Proof.

$$\begin{aligned}
\|u+v\|^2 + \|u-v\|^2 &= (u+v, u+v) + (u-v, u-v) = (u, u+v) + (v, u+v) + (u, u-v) - (v, u-v) \\
&= (u,u) + (u,v) + (v,u) + (v,v) + (u,u) - (u,v) - (v,u) + (v,v) = 2\|u\|^2 + 2\|v\|^2.
\end{aligned}$$

□

8 Spectral Theory of Self-Adjoint Mappings of a Euclidean Space into Itself

The following result will help us understand some details in the proof of Theorem 4′ (page 108, "It follows from this easily that we may choose an orthonormal basis consisting of real eigenvectors in each eigenspace N_a.")

Proposition 4. *Let X be a conjugate invariant subspace of $\mathbb{C}^n$ (i.e. X is invariant under conjugate operation). Then we can find a basis of X consisting of real vectors.*

Proof. We work by induction. First, assume $\dim X = 1$. $\forall v \in X$ with $v \neq 0$, we must have $\mathrm{Re}v \in X$ and $\mathrm{Im}v \in X$. At least one of them is non-zero and can be taken as a basis. Suppose for all conjugate invariant subspaces with dimension no more than k the claim is true. Let $\dim X = k+1$. $\forall v \in X$ with $v \neq 0$. If $\mathrm{Re}v$ and $\mathrm{Im}v$ are (complex) linearly dependent, there must exist $c \in \mathbb{C}$ and $v_0 \in \mathbb{R}^n$ such that $v = cv_0$, and we let $Y = \mathrm{span}\{v_0\}$; if $\mathrm{Re}v$ and $\mathrm{Im}v$ are (complex) linearly independent, we let $Y = \mathrm{span}\{v, \bar{v}\} = \mathrm{span}\{\mathrm{Re}v, \mathrm{Im}v\}$. In either case, Y is conjugate invariant. Let $Y^\perp = \{x \in X : \sum_{i=1}^n x_i \bar{y}_i = 0, \forall y = (y_1, \cdots, y_n)' \in Y\}$. Then clearly, $X = Y \bigoplus Y^\perp$ and $Y^\perp$ is also conjugate invariant. By assumption, we can choose a basis of $Y^\perp$ consisting exclusively of real vectors. Combined with the real basis of Y, we get a real basis of X. □

1.

Proof.

$$\left(x, \frac{M+M^*}{2}x\right) = \frac{1}{2}[(x, Mx) + (x, M^*x)] = \frac{1}{2}[(x, Mx) + (Mx, x)] = \frac{1}{2}[(x, Mx) + \overline{(x, Mx)}] = \mathrm{Re}(x, Mx).$$

□

3.

Proof. We prove $p_+ + p_0 = \max_{q(S)\geq 0} \dim S$. $p_- + p_0 = \max_{q(S)\leq 0} \dim S$ can be proved similarly. We shall use representation (11) for q in terms of the coordinates $z_1, \cdots, z_n$; suppose we label them so that $d_1, \cdots, d_p$ are nonnegative where $p = p_+ + p_0$, and the rest are negative. Define the subspace S_1 to consist of all vectors for which $z_{p+1} = \cdots = z_n = 0$. Clearly, $\dim S_1 = p$ and q is nonnegative on S_1. This shows $p_+ + p_0 = p \leq \max_{q(S)\geq 0} \dim S$. If $<$ holds, there must exist a subspace S_2 such that $q(S_2) \geq 0$ and $\dim S_2 > p = p_+ + p_0$. Define $P : S_2 \to S_3 := \{z : z_{p+1} = z_{p+2} = \cdots = z_n = 0\}$ by $P(z) = (z_1, \cdot, z_p, 0, \cdots, 0)$. Since $\dim S_2 > p = \dim S_3$, there exists some $z \in S_2$ such that $z \neq 0$ and $P(z) = 0$. This implies $z_1 = \cdots = z_p = 0$. So $q(z) = \sum_{i=1}^{p} d_i z_i^2 + \sum_{i=p+1}^{n} d_i z_i^2 = \sum_{i=p+1}^{n} d_i z_i^2 < 0$, contradiction. Therefore, our assumption is not true and $<$ cannot hold. □

4.

Proof. Write M in the column form $M = [c_1, \cdots, c_n]$ and multiply M to both sides for formula (24)', we get

$$HM = [Hc_1, \cdots, Hc_n] = MD = [c_1, \cdots, c_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1 c_1, \cdots, \lambda_n c_n],$$

where $\lambda_1, \cdots, \lambda_n$ are eigenvalues of M, including multiplicity. So we have $Hc_i = \lambda_i c_i$, $i = 1, \cdots, n$. This shows the columns of M are eigenvectors of H. □

5.

Proof. The essence of the generalization can be summarized as follows: $\langle x, y\rangle = (x, My)$ is an inner product and $M^{-1}H$ is self-adjoint under this new inner product, hence all the previous results apply.

Indeed, $\langle x, y\rangle$ is a bilinear function of x and y; it is symmetric since M is self-adjoint; and it is positive since M is positive. Combined, we can conclude $\langle x, y\rangle$ is an inner product.

Because M is positive, $Mx = 0$ has a unique solution 0. So M^{-1} exists. Define $U = M^{-1}H$ and we check U is self-adjoint under the new inner product $\langle\cdot,\cdot\rangle$. Indeed, $\forall x, y \in X$,

$$\langle Ux, y\rangle = (Ux, My) = (M^{-1}Hx, My) = (Hx, y) = (x, Hy) = (x, MM^{-1}Hy) = (x, MUy) = \langle x, Uy\rangle.$$

Applying the second proof of Theorem 4, with $(\cdot,\cdot)$ replaced by $\langle\cdot,\cdot\rangle$ and H replaced by $M^{-1}H$, we can verify claims (a)-(d) are true. □

6.

Proof. This is just Theorem 4 with $(\cdot,\cdot)$ replaced by $\langle\cdot,\cdot\rangle$ and H replaced by $M^{-1}H$, where $\langle\cdot,\cdot\rangle$ is defined in the solution of Exercise 5. □

7.

Proof. This is just Theorem 11 with $(\cdot,\cdot)$ replaced by $\langle\cdot,\cdot\rangle$ and H replaced by $M^{-1}H$, where $\langle\cdot,\cdot\rangle$ is defined in the solution of Exercise 5. □

8.

Proof. Under the new inner product $\langle\cdot,\cdot\rangle = (\cdot, M\cdot)$, $U = M^{-1}H$ is selfadjoint. By Theorem 4, all the eigenvalues of $M^{-1}H$ are real. If H is positive and $M^{-1}Hx = \lambda x$, then $\lambda\langle x, x\rangle = \langle x, M^{-1}Hx\rangle = (x, Hx) > 0$ for $x \neq 0$, which implies $\lambda > 0$. So under the condition that H is positive, all eigenvalues of $M^{-1}H$ are positive. □

10.

Proof. By Theorem 8, we can assume N has an orthonormal basis $v_1, \cdots, v_n$ consisting of genuine eigenvectors. We assume the eigenvalue corresponding to v_j is n_j. Then by letting $x = v_j$, $j = 1, \cdots, n$ and by the definition of $\|N\|$, we can conclude $\|N\| \geq \max|n_j|$. Meanwhile, $\forall x \in X$ with $\|x\| = 1$, there exist $a_1, \cdots, a_n \in \mathbb{C}$, so that $\sum |a_j|^2 = 1$ and $x = \sum a_j v_j$. So

$$\frac{\|Nx\|}{\|x\|} = \|\sum a_j n_j v_j\| = \sqrt{\sum |a_j n_j|^2} \leq \max_{1\leq j\leq n} |n_j| \sqrt{\sum |a_j|^2} = \max|n_j|.$$

This implies $\|N\| \leq \max|n_j|$. Combined, we can conclude $\|N\| = \max|n_j|$. □

Remark 12. *Compare the above result with formula (48) and Theorem 18 of Chapter 7.*

11.

Proof. $|S(a_1, \cdots, a_n)| = |(a_n, a_1, \cdots, a_{n-1})| = |(a_1, \cdots, a_n)|$. So S is an isometry in the Euclidean norm. To determine the eigenvalues and eigenvectors of S, note under the canonical basis $e_1, \cdots, e_n$, S corresponds to the matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

whose characteristic polynomial is $p(s) = |A - sI| = (-s)^n + (-1)^{n+1}$. So the eigenvalues of S are the solutions to the equation $s^n = 1$, i.e. $\lambda_k = e^{\frac{2\pi k}{n} i}$, $k = 1, \cdots, n$. Solve the equation $Sx_k = \lambda_k x_k$, we can obtain the general solution as $x_k = (\lambda_k^{n-1}, \lambda_k^{n-2}, \cdots, \lambda_k, 1)'$. After normalization, we have $x_k = \frac{1}{\sqrt{n}}(\lambda_k^{n-1}, \lambda_k^{n-2}, \cdots, \lambda_k, 1)'$. Therefore, for $i \neq j$,

$$(x_i, x_j) = \frac{1}{n}\sum_{k=1}^{n} \lambda_i^{k-1}\bar{\lambda}_j^{k-1} = \frac{1}{n}\sum_{k=1}^{n}(\lambda_i\bar{\lambda}_j)^{k-1} = \frac{1}{n}\frac{1-(\lambda_i\bar{\lambda}_j)^n}{1-\lambda_i\bar{\lambda}_j} = 0.$$

□

12. (i)

Proof. $A^*A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 13 \end{bmatrix}$, which has eigenvalues $7 \pm \sqrt{40}$. By Theorem 13, $\|A\| = \sqrt{7 + \sqrt{40}} \approx 3.65$. □

(ii)

Proof. This is consistent with the estimate obtained in Exercise 19 of Chapter 7: $3 \leq \|A\| \leq 3.7417$. □

13.

Proof. $\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 6 & -1 \\ 6 & 9 & 0 \\ -1 & 0 & 1 \end{bmatrix}$, which has eigenvalues 0, 1.6477, and 13.3523 By Theorem 13, the norm of the matrix is approximately $\sqrt{13.3523} \approx 3.65$. □

9 Calculus of Vector- and Matrix- Valued Functions

1.

Proof. Following the hint, note $\frac{d}{dt}(x(t),y) = (\dot{x}(t),y)+(x(t),\dot{y}) = 0$. So $(x(t),y)$ is a constant by fundamental lemma for scalar valued functions. Therefore $(x(t)-x(0),y)=0$, $\forall y \in \mathbb{K}^n$. This implies $x(t) \equiv x(0)$. □

2.

Proof. $A^{-1}(t)A(t) = I$. So $0 = \frac{d}{dt}\left[A^{-1}(t)A(t)\right] = \frac{d}{dt}A^{-1}(t)\cdot A(t) + A^{-1}(t)\dot{A}(t)$ and $\frac{d}{dt}A^{-1}(t) = \frac{d}{dt}A^{-1}(t)\cdot A(t)A^{-1}(t) = -A^{-1}(t)\dot{A}(t)A^{-1}(t)$. □

3.

Proof. $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. So

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{cases} I_{2\times 2} & \text{if } n \text{ is even} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \text{if } n \text{ is odd.} \end{cases}$$

Therefore, we have

$$\exp\{A+B\} = \sum_{n=0}^{\infty}\frac{1}{n!}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^n = \sum_{k=0}^{\infty}\frac{I_{2\times 2}}{(2k)!} + \sum_{k=0}^{\infty}\frac{1}{(2k+1)!}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{e+e^{-1}}{2}I_{2\times 2} + \frac{e-e^{-1}}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

□

4.

Proof. For any $\varepsilon > 0$, there exists $M > 0$, so that $\forall m \geq M$, $\sup_t \|\dot{E}_m(t) - F(t)\| < \varepsilon$. So $\forall m \geq M$, $\forall t, h$,

$$\begin{aligned}
& \left\|\frac{1}{h}[E_m(t+h) - E_m(t)] - F(t)\right\| \\
= & \left\|\frac{1}{h}\int_t^{t+h}[\dot{E}_m(s) - F(s)]ds + \frac{1}{h}\int_t^{t+h}F(s)ds - F(t)\right\| \\
\leq & \frac{\int_t^{t+h}\|\dot{E}_m(s) - F(s)\|ds}{h} + \left\|\frac{1}{h}\int_t^{t+h}F(s)ds - F(t)\right\| \\
< & \varepsilon + \left\|\frac{1}{h}\int_t^{t+h}F(s)ds - F(t)\right\|.
\end{aligned}$$

Under the assumption that F is continuous, we have

$$\begin{aligned}
\lim_{h\to 0}\left\|\frac{1}{h}[E(t+h) - E(t)] - F(t)\right\| &= \lim_{h\to 0}\lim_{m\to\infty}\left\|\frac{1}{h}[E_m(t+h) - E_m(t)] - F(t)\right\| \\
&\leq \varepsilon + \lim_{h\to 0}\left\|\frac{1}{h}\int_t^{t+h}F(s)ds - F(t)\right\| = \varepsilon.
\end{aligned}$$

Since ε is arbitrary, we must have $\lim_{h\to 0}\frac{1}{h}[E(t+h) - E(t)] = F(t)$. □

5.

Proof. By formula (12), $\dot{E}_m(t) = \sum_{k=1}^{m} \sum_{i=0}^{k-1} \frac{1}{k!} A^i(t)\dot{A}(t)A^{k-i-1}(t)$. So for m and n with $m < n$,

$$\begin{aligned}\|\dot{E}_m(t) - \dot{E}_n(t)\| &\leq \sum_{k=m+1}^{n} \sum_{i=0}^{k-1} \frac{\|A^i(t)\dot{A}(t)A^{k-i-1}(t)\|}{k!} = \sum_{k=m+1}^{n} \sum_{i=0}^{k-1} \frac{\|A(t)\|^{k-1}\|\dot{A}(t)\|}{k!} \\ &= \sum_{k=m+1}^{n} \frac{\|A(t)\|^{k-1}}{(k-1)!}\|\dot{A}(t)\| = \|\dot{A}(t)\|[e_n(\|A(t)\|) - e_m(\|A(t)\|)] \to 0\end{aligned}$$

as $m, n \to \infty$. This shows $(\dot{E}_m(t))_{m=1}^{\infty}$ is a Cauchy sequence, hence convergent. □

6.

Proof. Apply formula (10) to $Y(t) = e^{At}$, we have $\frac{d}{dt}\log\det Y(t) = \mathrm{tr}(e^{-At}e^{At}A) = \mathrm{tr}A$. Integrating from 0 to t, we get $\log\det Y(t) - \log\det Y(0) = t\mathrm{tr}A$. So $\det Y(t) = e^{t\mathrm{tr}A}$. In particular, $\det e^A = e^{\mathrm{tr}A}$. □

7.

Proof. Without loss of generality, we can assume A is a Jordan matrix. Then e^A is an upper triangular matrix and its entries on the diagonal line have the form e^a, where a is an eigenvalue of A. So all eigenvalues of e^A are the exponentials of eigenvalues of A. □

8. (a)

Proof. The total number of "free" entries is $\frac{n(n+1)}{2}$. The entries on the diagonal line must be real. So the dimension is $\frac{n(n+1)}{2} \times 2 - n = n^2$. □

(b)

Proof. Similar to the argument in the text, the total number of complex parameters that determine the eigenvectors is $(n-1) + \cdots + 2 = \frac{n(n-1)}{2} - 1$. This is equivalent to $n(n-1) - 2$ real parameters. The number of distinct (real) eigenvalues is $n - 1$. So the dimension $= n^2 - n - 2 + n - 1 = n^2 - 3$. □

9.

Proof. See the Matlab program **aoc.m**.

```
function aoc

%AOC illustrates the avoidance-of-crossing phenomenon
%    of the neighboring eigenvalues of a continuous
%    symmetric matrix. This is Exercise 9, Chapter 9
%    of the textbook, Linear Algebra and Its Applications,
%    2nd Edition, by Peter Lax.

% Initialize global variables
matrixSize = 10;
lowerBound = 0.01; %lower bound of t's range
upperBound = 3; %upper bound of t's range
stepSize = 0.1;
t = lowerBound:stepSize:upperBound;

% Generate random symmetric matrix
temp1 = rand(matrixSize);
temp2 = rand(matrixSize);
M = temp1+temp1';
B = temp2+temp2';
```

```
% Initialize eigenvalue matrix to zeros;
% use each column to store eigenvalues for
% a given parameter
eigenval = zeros(matrixSize,numel(t));
for i = 1:numel(t)
    eigenval(:,i) = eig(B+t(i)*M);
end

% Plot eigenvalues according to values of parameter
hold off;
disp(['There are ', num2str(matrixSize), ' eigenvalue curves.']);
disp(' ');
for j = 1:matrixSize
    disp(['Eigenvalue curve No. ', num2str(j),'. Press ENTER to continue...']);
    plot(t, eigenval(j,:));
    xlabel('t');
    ylabel('eigenvalues');
    title('Matlab illustration of Avoidance of Crossing');
    hold on;
    pause;
end
hold off;
```

□

10 Matrix Inequalities

1.

Proof. Suppose H has k distinct eigenvalues $\lambda_1, \cdots, \lambda_k$. Denote by X_j the subspace consisting of eigenvectors of H pertaining to the eigenvalue λ_j. Then H can be represented as $H = \sum_{j=1}^k \lambda_j P_j$ were P_j is the projection to X_j. Let A be a positive square root of H, we claim A has to be $\sum_{j=1}^k \sqrt{\lambda_j} P_j$. Indeed, if α is an eigenvalue of A and x is an eigenvector of A pertaining to α, then α^2 is an eigenvalue of H and x is an eigenvector of H pertaining to α^2. So we can assume A has m distinct eigenvalues $\alpha_1, \cdots, \alpha_m$ ($m \le k$) with $\alpha_i = \sqrt{\lambda_i}$ ($1 \le i \le m$). Denote by Y_i the subspace consisting of eigenvectors of A pertaining to α_i. Then $Y_i \subset X_i$. Since $H = \bigoplus_{i=1}^m Y_i = \bigoplus_{j=1}^k X_j$, we must have $m = k$ and $Y_i = X_i$, otherwise at lease one of the $\le$ in the sequence of inequalities $\dim H = \sum_{i=1}^m \dim Y_i \le \sum_{i=1}^m \dim X_i \le \sum_{i=1}^k \dim X_i = \dim H$ is $<$, contradiction. So A can be uniquely represented as $A = \sum_{j=1}^k \sqrt{\lambda_j} P_j$, the same as $\sqrt{H}$ defined in formula (6). □

2.

Proposition 5. *(i) The identity I is nonnegative. (ii) If M and N are nonnegative, so is their sum $M + N$, as well as aM for any nonnegative number a. (iii) If H is nonnegative and Q is invertible, we have $Q^*HQ \ge 0$. (iv) H is nonnegative if and only if all its eigenvalues are nonnegative. (vi) Every nonnegative mapping has a nonnegative square root, uniquely determined.*

Proof. (i) and (ii) are obvious. For part (iii), we write the quadratic form associated with Q^*HQ as

$$(x, Q^*HQx) = (Qx, HQx) = (y, Hy) \ge 0,$$

where $y = Qx$. For part (iv), by the selfadjointness of H, there exists an orthogonal basis of eigenvectors. Denote these by h_j and the corresponding eigenvalue by a_j. Then any vector x can be expressed as a linear combination of the h_j's: $x = \sum_j x_j h_j$. So $(x, Hx) = \sum_{i,j} (x_i h_i, x_j a_j h_j) = \sum_{j=1}^n a_j |x_j|^2$. From the formula

it is clear that $(x, Hx) \geq 0$ for any x if and only if $a_j \geq 0$, $\forall j$. For part (vi), the proof is similar to that of positive mappings and we omit the lengthy proof. Cf. also solution to Exercise 10.1. $\square$

3.

Proof. Let A be a mapping that maps the vector $(0,1)'$ to $(0,\alpha_2)'$ with $\alpha_2 > 0$ sufficiently small and $(1,0)'$ to $(\alpha_1, 0)'$ with $\alpha_1 > 0$ sufficiently large. Let B be a mapping that maps the vector $(1,1)'$ to $(\lambda_1, \lambda_1)'$ with $\lambda_1 > 0$ sufficiently small and $(-1,1)'$ to $(-\lambda_2, \lambda_2)'$ with $\lambda_2 > 0$ sufficiently large. Then both A and B are positive mappings, and we can find x between $(1,1)'$ and $(0,1)'$ so that $(Ax, Bx) < 0$. By the analysis in the paragraph below formula $(14)'$, $AB + BA$ is not positive. More precisely, we have $A = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}$ and $B = \frac{1}{2}\begin{pmatrix} \lambda_1 + \lambda_2 & \lambda_1 - \lambda_2 \\ \lambda_1 - \lambda_2 & \lambda_1 + \lambda_2 \end{pmatrix}$. $\square$

4.

Proof. By Theorem 5 and induction, it is easy to prove (a) and (b). For (c), we follow the hint. If M has the spectral resolution $M = \sum_{i=1}^k \lambda_i P_i$, $\log M$ is defined as

$$\log M = \sum_{i=1}^k \log \lambda_i P_i = \sum_{i=1}^k \lim_{m\to\infty} m(\lambda_i^{\frac{1}{m}} - 1)P_i = \lim_{m\to\infty} m\left(\sum_{i=1}^k \lambda_i^{\frac{1}{m}} P_i - \sum_{i=1}^k P_i\right) = \lim_{m\to\infty} m(M^{\frac{1}{m}} - I).$$

So $\log M = \lim_{m\to\infty} m(M^{\frac{1}{m}} - I) \leq \lim_{m\to\infty} m(N^{\frac{1}{m}} - I) = \log N$. $\square$

5.

Proof. (from the textbook's solution, pp. 291) Choose A and B as in Exercise 3, that is positive matrices whose symmetrized product is not positive. Set

$$M = A, N = A + tB,$$

t sufficiently small positive number. Clearly, $M < N$.

$$N^2 = A^2 + t(AB + BA) + t^2 B^2;$$

for t small the term $t^2 B$ is negligible compared with the linear term. Therefore for t small N^2 is not greater than M^2. $\square$

6.

Proof. For $f(z) = az + b - \int_0^\infty \frac{dm(t)}{z+t}$, we have

$$f(z + \Delta z) - f(z) = a\Delta z + \Delta z \cdot \int_0^\infty \frac{dm(t)}{(z + \Delta z + t)(z+t)}.$$

So if we can show $\lim_{\Delta z \to 0} \int_0^\infty \frac{dm(t)}{(z+\Delta z+t)(z+t)}$ exists and is finite, $f(z)$ is analytic by definition. Indeed, if $\mathrm{Im} z > 0$, for Δz sufficiently small, we have

$$\left|\frac{1}{z + \Delta z + t}\right| \leq \frac{1}{|z+t| - |\Delta z|} \leq \frac{1}{\mathrm{Im} z - |\Delta z|} \leq \frac{2}{\mathrm{Im} z}.$$

So by Dominated Convergence Theorem, $\lim_{\Delta z\to 0} \int_0^\infty \frac{dm(t)}{(z+\Delta z+t)(z+t)}$ exists and is equal to $\int_0^\infty \frac{dm(t)}{(z+t)^2}$, which is finite. To see $\mathrm{Im} f(z) > 0$ for $\mathrm{Im} z > 0$, we note

$$\mathrm{Im} f(z) = a\mathrm{Im} z - \mathrm{Im} \int_0^\infty \frac{dm(t)}{\mathrm{Re} z + t + i\mathrm{Im} z} = \mathrm{Im} z \left[a + \int_0^\infty \frac{dm(t)}{(\mathrm{Re} z + t)^2 + (\mathrm{Im} z)^2}\right].$$

$\square$

7.

Proof. Consider the Euclidean space $L^2(-\infty, 1]$, with the inner product $(f,g) := \int_{-\infty}^1 f(t)g(t)dt$. Choose $f_j = e^{r_j(t-1)}$, $j = 1, \cdots, m$, then the associated Gram matrix is

$$G_{ij} = (f_i, f_j) = \int_{-\infty}^1 \frac{e^{(r_i+r_j)t}}{e^{r_i+r_j}} dt = \frac{1}{r_i + r_j}.$$

Clearly, $(f_j)_{j=1}^m$ are linearly independent. So G is positive. $\square$

8.

Proof. We apply the change of variable formula as follows

$$\int_{-\infty}^{\infty} e^{-z^2} dz = \sqrt{\int_{\mathbb{R}^2} e^{-x^2-y^2} dx dy} = \sqrt{\int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2} r dr} = \sqrt{2\pi \cdot \frac{1}{2}} = \sqrt{\pi}.$$

$\square$

9.

Proof. The extension is straightforward, just replace the paragraph (on page 161) "If S is a subspace of V, then $T = S$ and $\dim T = \dim S$. ... It follows that

$$\dim S - 1 \leq \dim T$$

as asserted." with the following one: Let $T = S \cap V$ and $T_1 = S \cap V^\perp$, where $V^\perp$ stands for the compliment of V in U. Then $\dim T + \dim T_1 = \dim S$. Since $\dim T_1 \leq \dim V^\perp = n - (n - m) = m$, $\dim T \geq \dim S - m$.

The rest of the proof is the same as the proof of Theorem 14 and we can conclude that

$$p_\pm(A) - m \leq p_\pm(B) \leq p_\pm(A).$$

$\square$

10.

Proof. For any x, $(x, (N - M - dI)x) = (x, (N - M)x) - d\|x\|^2 \leq \|N - M\|\|x\|^2 - d\|x\|^2 = 0$. Similarly, $(x, (M - N - dI)x) = (x, (M - N)x) - d\|x\|^2 \leq \|M - N\|\|x\|^2 - d\|x\|^2 \leq 0$. $\square$

11.

Proof. It's easy to see the problem can be reduced to the case $k = 2$. To prove this case, we note if $m_1 \leq m_2$ and $n_{p_1} \geq n_{p_2}$, we have

$$m_2 n_{p_1} + m_1 n_{p_2} - m_2 n_{p_2} - m_1 n_{p_1} = (m_2 - m_1)(n_{p_1} - n_{p_2}) \geq 0.$$

$\square$

12.

Proof. Assume Z is not invertible. Then there exists $x \neq 0$ such that $Zx = 0$. In particular, this implies $(x, Zx) = (x, Z^*x) = 0$. Sum up these two, we get $(x, (Z + Z^*)x) = 0$. Contradictory to the assumption that the selfadjoint part of Z is positive. For any $x \neq 0$, there exists $y \neq 0$ so that $x = Zy$. So

$$(x, (Z^{-1}+(Z^{-1})^*)x) = (x, Z^{-1}x)+(x, (Z^{-1})^*x) = (Zy, y)+(Z^{-1}x, x) = (y, Z^*y)+(y, Zy) = (y, (Z+Z^*)y) > 0.$$

This shows the selfadjoint part of Z^{-1} is positive. $\square$

13.

Proof. Exercise Problem 14 has proved the claim for non-zero eigenvalues. Since the dimensions of the spaces of generalized eigenvectors of AA^* and A^*A are both equal to the dimension of the underlying Euclidean space, we conclude by Spectral Theorem that their zero eigenvalues must have the same multiplicity. $\square$

14.

Proof. Suppose a is a non-zero eigenvalue of AA^* and x is an eigenvector of AA^* pertaining to a: $AA^*x = ax$. Applying A^* to both sides, we get $A^*A(A^*x) = aA^*x$. Since $a \neq 0$ and $x \neq 0$, $A^*x \neq 0$ by $AA^*x = ax$. Therefore, a is an eigenvalue of A^*A with A^*x an eigenvector of A^*A pertaining to a. By symmetry, we conclude AA^* and A^*A have the same set of non-zero eigenvalues.

Fix a non-zero eigenvalue a, and suppose $x_1, \cdots, x_m$ is a basis for the space of generalized eigenvectors of AA^* pertaining to a. Since $a \neq 0$, we can claim $A^*x_1, \cdots, A^*x_m$ are linearly independent. Indeed, assume not, then there must exist $\alpha_1, \cdots, \alpha_m$ not all equal to 0, such that $\sum_{i=1}^m \alpha_i A^*x_i = 0$. This implies $a(\sum_{i=1}^m \alpha_i x_i) = \sum_{i=1}^m \alpha_i AA^*x_i = A(\sum_{i=1}^m \alpha_i A^*x_i) = 0$, which further implies $x_1, \cdots, x_m$ are linearly dependent since $a \neq 0$. Contradiction.

This shows the dimension of the space of generalized eigenvectors of AA^* pertaining to a is no greater than that of the space of generalized eigenvectors of A^*A pertaining to a. By symmetry, we conclude the spaces of generalized eigenvectors of AA^* and A^*A pertaining to the same nonzero eigenvalue have the same dimension. Combined, we can conclude AA^* and A^*A have the same non-zero eigenvalues with the same (algebraic) multiplicity. $\square$

Remark 13. *The multiplicity referred to in this problem is understood as algebraic multiplicity, which is equal to the dimension of the space of generalized eigenvectors.*

15.

Proof. Let $Z = \begin{pmatrix} 1+bi & 3 \\ 0 & 1+bi \end{pmatrix}$ where b could be any real number. Then the eigenvalue of Z, $1+bi$, has positive real part. Meanwhile, $Z + Z^* = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$ has characteristic polynomial $p(s) = (2-s)^2 - 9 = (s-5)(s+1)$. So $Z + Z^*$ has eigenvalue 5 and -1, and therefore cannot be positive. $\square$

16.

Proof. Suppose A and B are selfadjoint. Then for any x and y,

$$(x, (AB-BA)^*y) = ((AB-BA)x, y) = (ABx, y) - (BAx, y) = (x, BAy) - (x, ABy) = (x, -(AB-BA)y).$$

So $(AB-BA)^* = -(AB-BA)$. $\square$

11 Kinematics and Dynamics

1.

Proof. We note $\frac{d}{dt}(M(t)M^*(t)) = \dot{M}(t)M^*(t) + M(t)\dot{M}^*(t) = A(t) + A^*(t) = 0$. So $M(t)M^*(t) \equiv M(0)M^*(0) = I$. Also, $f(t) = \det M(t)$ is continuous function of t and takes values either 1 or -1 by the isometry property of $M(t)$. Since $f(0) = 1$, we have $f(t) \equiv 1$. By Theorem 1, $M(t)$ is a rotation for every t. $\square$

2.

Proof. $\lim_{h\to 0} \frac{M(t+h)-M(t)}{h} = \lim_{h\to 0} \frac{e^{tA}(e^{hA}-I)}{h} = Ae^{tA}$, i.e. $\dot{M}(t) = AM(t)$. Clearly $M(0) = I$. $\square$

3.

Proof. The reason we need commutativity is that the following equation is required in the calculation of derivative:

$$\begin{aligned}\frac{1}{h}(M(t+h)-M(t)) &= \frac{1}{h}\left(e^{\int_0^{t+h}A(s)ds}-e^{\int_0^t A(s)ds}\right)\\ &= \frac{1}{h}\left(e^{\int_0^t A(s)ds+\int_t^{t+h}A(s)ds}-A^{\int_0^t A(s)ds}\right)\\ &= \frac{1}{h}e^{\int_0^t A(s)ds}\left(e^{\int_t^{t+h}A(s)ds}-I\right),\end{aligned}$$

i.e. $e^{\int_0^t A(s)ds+\int_t^{t+h}A(s)ds}=e^{\int_0^t A(s)ds}e^{\int_t^{t+h}A(s)ds}$. So when this commutativity holds,

$$\dot{M}(t)=\lim_{h\to 0}\frac{1}{h}(M(t+h)-M(t))=M(t)A(t).$$

□

4.

Proof. If $f=(x,y,z)^T$ satisfies $Af=0$, we must have

$$\begin{cases}ay+bz=0\\ -ax+cz=0\\ -bx-cy=0.\end{cases}$$

By discussing various possibilities ($a,b,c=0$ or not), we can check f is a multiple of $(-c,b,-a)^T$. □

5.

Proof.

$$\det(sI-A)=\det\begin{pmatrix}\lambda & -a & -b\\ a & \lambda & -c\\ b & c & \lambda\end{pmatrix}=\lambda(\lambda^2+c^2)-a(-a\lambda+bc)+b(ac+b\lambda)=\lambda^3+\lambda(c^2+b^2+a^2).$$

Solving it gives us the other two eigenvalues. □

6.

Proof. Since $A=\begin{pmatrix}0 & a & b\\ -a & 0 & c\\ -b & -c & 0\end{pmatrix}$ is anti-symmetric, $M(t)M^*(t)=e^{tA}e^{tA^*}=e^{tA}e^{-tA}=I$. By Exercise 7 of Chapter 9, all eigenvalues of e^{At} has the form of e^{at}, where a is an eigenvalue of A. Since the eigenvalues of A are 0 and $\pm ik$ with $k=\sqrt{a^2+b^2+c^2}$ (Exercise 5), the eigenvalues of e^{At} are 1 and $e^{\pm ikt}$. This implies $\det e^{At}=1\cdot e^{ikt}\cdot e^{-ikt}=1$. By Theorem 1, $M=e^{At}$ is a rotation. Let f be given by formula (16). From $Af=0$ we deduce that $e^{At}f=f$; thus f is the axis of the rotation e^{At}. The trace of e^{At} is $1+e^{ikt}+e^{-ikt}=2\cos kt+1$. According to formula $(4)'$, the angle of rotation θ of e^{At} satisfies $2\cos\theta+1=\mathrm{tr}e^{At}$. This shows that $\theta=kt=\sqrt{a^2+b^2+c^2}$. □

7.

Proof. $(AB-BA)^*=(AB)^*-(BA)^*=B^*A^*-A^*B^*=(-B)(-A)-(-A)(-B)=BA-AB=-(AB-BA)$. □

8.

Proof. See the solution in the textbook, page 294. □

9.

Proof. See the solution in the textbook, page 294. □

10.

Proof. It suffices to note that the set of 2n functions, $\{(\cos c_j t)h_j, (\sin c_j t)h_j\}_{j=1}^n$, are linearly independent, since any two of them are orthogonal when their subscripts are distinct. □

12 Convexity

The following results will help us understand some details in the proofs of Theorem 6 and Theorem 10.

Proposition 6. *Let S be an arbitrary subset of X and x an interior point of S. For any real linear function l defined on X, if $l \not\equiv 0$, then $l(x)$ is an interior point of $\Gamma = l(S)$ in the topological sense.*

Proof. We can find $y \in X$ so that $l(y) \neq 0$. Then for t sufficiently small, $l(x) + tl(y) = l(x + ty) \in \Gamma$. So Γ contains an interval which contains $l(x)$, i.e. $l(x)$ is an interior point of Γ under the topology of $\mathbb{R}^1$. □

Corollary 1. *If K is an open convex set and l is a linear function with $l \not\equiv 0$, $\Gamma = l(K)$ is an open interval.*

Proof. Note Γ is convex and open in $\mathbb{R}^1$ in the topological sense. □

Proposition 7. *Let K be a convex set and K_0 the set of all interior points of K. Then K_0 is convex and open.*

Proof. (Convexity) $\forall x, y \in K_0$ and $a \in [0, 1]$. For any $z \in X$, $[ax+(1-a)y]+tz = a(x+tz)+(1-a)(y+tz) \in K$ when t is sufficiently small, since x, y are interior points of K and K is convex.

(Openness) Fix $x \in K_0$, $\forall y_1 \in X$. We need to show for t sufficiently small, $x+ty_1 \in K_0$. Indeed, $\forall y_2 \in X$, we can find a common $\varepsilon > 0$, so that whenever $(t_1, t_2) \in [-\varepsilon, \varepsilon] \times [-\varepsilon, \varepsilon]$, $x + t_1 y_1 \in K$ and $x + t_2 y_2 \in K$. Fix any $t^* \in [-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$, by the convexity of K, $x + t^* y_1 + t^{**} y_2 = \frac{1}{2}(x + 2t^* y_1) + \frac{1}{2}(x + 2t^{**} y_2) \in K$ when $t^{**} \in [-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$. This shows $x + t^* y_1 \in K_0$. Since t^* is arbitrarily chosen from $[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$, we conclude for t sufficiently small, $x + ty_1 \in K_0$. That is, x is an interior point of K_0. By the arbitrariness of x, K_0 is open. □

1.

Proof. The verification is straightforward and we omit it. □

2.

Proof. These propositions are immediate consequences of the definition of convexity. □

3.

Proof. Fix a point $x \in \{z : l(z) < c\}$. For any $y \in X$, $f(t) = l(x + ty) = l(x) + tl(y)$ is a continuous function of t, with $f(0) = l(x) < c$. By continuity, $f(t) < c$ for t sufficiently small. So $x + ty \in \{z : l(z) < c\}$ for t sufficiently small, i.e. x is an interior point. Since x is arbitrarily chosen, we have proved $\{z : l(z) < c\}$ is open. □

4.

Proof. The convexity of $A + B$ is Theorem 1(b). To see the openness, $\forall x \in A$, $y \in B$. For any $z \in X$, $(x + y) + tz = (x + tz) + y$. For t sufficiently small, $x + tz \in A$. So $(x + y) + tz \in A + B$ for t sufficiently small. This shows $A + B$ is open. □

5.

Proof. That K is a convex set is trivial to see. It's also clear that $p(0) = 0$. For any $x \in \mathbb{R}^n \setminus \{0\}$, when $\varepsilon \in (0, a)$, $r_\varepsilon = \frac{\|x\|}{a-\varepsilon}$ satisfies $r_\varepsilon > 0$ and $x/r_\varepsilon \in K$. So $p(x) \le \frac{\|x\|}{a-\varepsilon}$. By letting $\varepsilon \to 0$, we conclude $p(x) \le \|x\|/a$. If " $<$ " holds, we can find $r > 0$ such that $r < \frac{\|x\|}{a}$ and $\frac{x}{r} \in K$. But $r < \frac{\|x\|}{a}$ implies $a < \frac{\|x\|}{r}$ and hence $\frac{x}{r} \notin K$. Contradiction. Combined, we conclude $p(x) = \frac{\|x\|}{a}$. □

6.

Proof. See the textbook's solution. □

7.

Proof. $\forall x, y \in K$, we have $p(x) < 1$ and $p(y) < 1$. For any $a \in [0, 1]$, $p(ax+(1-a)y) \le p(ax)+p((1-a)y) = ap(x) + (1-a)p(y) < a + (1-a) = 1$. This shows K is convex. To see K is open, fix $x \in K$ and choose any $z \in X$. Then $p(x + tz) \le p(x) + tp(z)$. So for t sufficiently small such that $tp(z) < 1 - p(x)$, we have $p(x + tz) \le p(x) + tp(z) < p(x) + 1 - p(x) = 1$, i.e. $x + tz \in K$. This shows K is open. □

8.

Proof. $\forall \varepsilon > 0$, there exists $x(\varepsilon) \in S$, so that

$$q_S(m + l) = \sup_{x \in S}(l + m)(x) < (l + m)(x(\varepsilon)) + \varepsilon \le \sup_{x \in S} l(x) + \sup_{x \in S} m(x) + \varepsilon = q_S(m) + q_S(l) + \varepsilon.$$

By the arbitrariness of ε, we conclude $q_S(m + l) \le q_S(m) + q_S(l)$. □

9.

Proof. $q_{S+T}(l) = \sup_{x\in S, y\in T} l(x + y) = \sup_{x\in S, y\in T}[l(x) + l(y)] \le \sup_{x\in S, y\in T}[q_S(l) + q_T(l)] = q_S(l) + q_T(l)$. Conversely, $\forall \varepsilon > 0$, there exists $x_0 \in S$, $y_0 \in T$, s.t. $q_S(l) < l(x_0) + \frac{\varepsilon}{2}$, $q_T(l) < l(y_0) + \frac{\varepsilon}{2}$. So $q_S(l) + q_T(l) < l(x_0 + y_0) + \varepsilon \le q_{S+T}(l) + \varepsilon$. By the arbitrariness of ε, $q_S(l) + q_T(l) \le q_{S+T}(l)$. Combined, we get $q_{S+T}(l) = q_S(l) + q_T(l)$. □

10.

Proof. $q_{S\cup T}(l) = \sup_{x\in S\cup T} l(x) \ge \sup_{x\in S} l(x) = q_S(l)$. Similarly, $q_{S\cup T}(l) \ge q_T(l)$. Therefore, we have $q_{S\cup T}(l) \ge \max\{q_S(l), q_T(l)\}$. For any $\varepsilon > 0$ sufficiently small, we can find $x_\varepsilon \in S \cup T$, such that $q_{S\cup T}(l) \le l(x_\varepsilon) + \varepsilon$. But $l(x_\varepsilon) \le \max\{q_S(l), q_T(l)\}$. So $q_{S\cup T}(l) \le \max\{q_S(l), q_T(l)\} + \varepsilon$. Let $\varepsilon \to 0$, we get $q_{S\cup T}(l) \le \max\{q_S(l), q_T(l)\}$. Combined, we can conclude $q_{S\cup T}(l) = \max\{q_S(l), q_T(l)\}$. □

11.

Proof. If for any $a \in (0, 1)$, $l(ax + (1-a)y) = al(x) + (1-a)l(y) \le c$, by continuity, we have $l(x) \le c$ and $l(y) \le c$. This shows $\{x : l(x) \le c\}$ is a closed convex set. □

12.

Proof. Convexity is obvious. For closedness, note $f(t) = \|tx + (1-t)y\|$ is a continuous function of t. So if $f(t) \le t$ for any $t \in (0, 1)$, $f(0) = \|y\| \le 1$ and $f(1) = \|x\| \le 1$. So the unit ball $B(0, 1)$ is closed. Combined, we conclude $B(0, 1) = \{x : \|x\| \le 1\}$ is a closed convex set. □

13.

Proof. Suppose H and K are closed convex sets. Theorem 1(a) says $H \cap K$ is also convex. Moreover, if for any $a \in (0, 1)$, $ax + (1-a)y \in H \cap K$, then the closedness of H and K implies $a, b \in H$ and $a, b \in K$, i.e. $a, b \in H \cap K$. So $H \cap K$ is closed. □

14.

Proof. **Proof of Theorem 7**: Suppose K has an interior point x_0. If a linear function l and a real number c determine a closed half-space that contains $K - x_0$ but not $y - x_0$, i.e. $l(x - x_0) \le c$, $\forall x \in K$ and $l(y - x_0) > c$, then l and $c + l(x_0)$ determine a closed half-space that contains K but not y, i.e. $l(x) \le c + l(x_0)$ and $l(y) > c + l(x_0)$. So without loss of generality, we can assume $x_0 = 0$. Note the convexity and closedness are preserved under translation, so this simplification is all right for this problem's purpose.

Define gauge function p_K as in (5). Then we can show $p_K(x) \le 1$ if and only if $x \in K$. Indeed, if $x \in K$, then $p_K(x) \le 1$ by deifinition. Conversely, if $p_K(x) \le 1$, then for any $\varepsilon > 0$, there exists $r(\varepsilon) < 1 + \varepsilon$ so that $\frac{x}{r(\varepsilon)} \in K$. We choose $r(\varepsilon) > 1$ and note $\frac{x}{r(\varepsilon)} = a(\varepsilon) \cdot 0 + (1 - a(\varepsilon)) \cdot x$ with $a(\varepsilon) = 1 - \frac{1}{r(\varepsilon)}$. As $r(\varepsilon)$ can be as close to 1 as we want when $\varepsilon \downarrow 0$, $a(\varepsilon)$ can be as close to 0 as we want. Meanwhile, 0 is an interior point of K, so for r large enough, $\frac{x}{r} \in K$. This shows for a close enough to 1, $a \cdot 0 + (1 - a) \cdot x \in K$. Combined, we conclude K contains the open segment $\{a \cdot 0 + (1 - a) \cdot x : 0 < a < 1\}$. By definition of closedness, $x \in K$. The rest of the proof is completely analogous to that of Theorem 3, with $p(x) < 1$ replaced by $p(x) \le 1$.

If K has no interior point, we have two possibilities. Case one, y and K are not on the same hyperplane. In this case, there exists a linear function l and a real number c, such that $l(x) = c(\forall x \in K)$ but $l(y) \ne c$. By considering $-l$ if necessary, we can have $l(x) = c(\forall x \in K)$ and $l(y) > c$. So the half-space $\{x : l(x) \le c\}$ contains K, but not y. Case two, y and K reside on the same hyperplane. Then the dimension of the ambient space for y and K can be reduced by 1. Work by induction and note the space is of finite dimension, we can finish the proof.

Proof of Theorem 8: By definition of (16), if $x \in K$, then $l(x) \le q_K(l)$. Conversely, suppose y is not in K, then there exists $l \in X'$ and a real number c such that $l(x) \le c$, $\forall x \in K$ and $l(y) > c$. This implies $l(y) > q_K(l)$. Combined, we conclude $x \in K$ if and only if $l(x) \le q_K(l)$, $\forall l \in X'$.

Remark: From the above solution and the proof of Theorem 3, we can see a useful routine for proving results on convex sets: first assume the convex set has an interior point and use the gauge function, which often helps to construct the desired linear functionals via Hahn-Banach Theorem. If there exists no interior point, reduce the dimension by 1 and work by induction. Such a use of interior points as the criterion for a dichotomy is also present in the proof of Theorem 10 (Carathéodory). □

15.

Proof. Denote by $\widehat{S}$ the closed convex hull of S, and define $\Gamma_l = \{x : l(x) \le q_S(l)\}$ where $l \in X'$. Then it is easy to see each Γ_l is a closed convex set containing S, so $\widehat{S} \subseteq \cap_{l \in X'} \Gamma_l$. For the other direction, suppose $\cap_{l \in X'} \Gamma_l \setminus \widehat{S} \ne \emptyset$ and we choose a point x from $\cap_{l \in X'} \Gamma_l \setminus \widehat{S}$. By Theorem 8, there exists $l_0 \in X'$ such that $l_0(x) > q_{\widehat{S}}(l_0) \ge q_S(l_0)$. So $x \notin \Gamma_{l_0}$, contradiction. Combined, we conclude $\widehat{S} = \cap_{l \in X'} \Gamma_l = \{x : l(x) \le q_S(l), \forall l \in X'\}$. □

16.

Proof. Suppose $\lambda_1, \cdots, \lambda_m$ satisfy $\lambda_1, \cdots, \lambda_m \in (0, 1)$ and $\sum_{i=1}^m \lambda_i = 1$. We need to show $\sum_{i=1}^m \lambda_i x_i \in K$, where K is the convex set to which $x_1, \cdots, x_m$ belong. Indeed, since $\sum_{i=1}^m \lambda_i x_i = (\lambda_1 + \cdots + \lambda_{m-1}) \sum_{i=1}^{m-1} \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_{m-1}} x_i + \lambda_m x_m$, it suffices to show $\sum_{i=1}^{m-1} \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_{m-1}} x_i \in K$. Working by induction, we are done. □

17.

Proof. Suppose x is an interior point of K. $\forall y \in X$, for t sufficiently small, $x + ty \in K$. In particular, we can find $\varepsilon > 0$ so that $x + \varepsilon y \in K$ and $x - \varepsilon y \in K$. Since x can be represented as $x = \frac{(x + \varepsilon y) + (x - \varepsilon y)}{2}$, we conclude x is not an extreme point. □

18.

Proof. Let S be a permutation matrix as defined in formula (25). Then clearly $S_{ij} \ge 0$. Furthermore, $\sum_{i=1}^n S_{ij} = \sum_{i=1}^n \delta_{p(i)j}$, where j is fixed and is equal to $p(i_0)$ for some i_0. So $\sum_{i=1}^n S_{ij} = 1$. Finally, $\sum_{j=1}^n S_{ij} = \sum_{j=1}^n \delta_{ip^{-1}(j)}$, where i is fixed and is equal to $p^{-1}(j_0)$ for some j_0. So $\sum_{j=1}^n S_{ij} = 1$. Combined, we conclcude S is a doubly stochastic matrix. □

19.

Proof. The textbook's solution demonstrates the case of dimension 3. Counterexamples for higher dimensions can be obtained by building permutation matrices upon the case of dimension 3. □

20. Suppose K is a convex subset of an n-dimension linear space X. We have the following properties.

(1) If x is an interior point of K in the linear sense, then x is an interior point of K in the topological sense. Consequently, being open in the linear sense is the same as being open in the topological sense.

Proof. Let $e_1, \cdots, e_n$ be a basis of X. There exists $\varepsilon > 0$ so that for any $t_i \in (-\varepsilon, \varepsilon)$, $x+t_ie_i \in K$, $i = 1, \cdots, n$. For any $y \in X$ which is close enough to x, the norm of $y - x$ can be very small so that if we write y as $y = x + \sum_{i=1}^n a_ie_i$, $|a_i| < \frac{\varepsilon}{n}$. Since for $t_i \in (-\frac{\varepsilon}{n}, \frac{\varepsilon}{n})$ $(i = 1, \cdots, n)$, $x + \sum_{i=1}^n t_ie_i = \sum_{i=1}^n \frac{1}{n}(x + nt_ie_i) \in K$ by the convexity of K, we conclude $y \in K$ if y is sufficiently close to x. This shows x is an interior point of K. □

(2) If K is closed in the linear sense, it is closed in the topological sense.

Proof. Suppose $(x_k)_{k=1}^\infty \subseteq K$ and $x_k \to x$ in the topological sense, we need to show $x \in K$. We work by induction. The case $n = 1$ is trivial, because x is necessarily an endpoint of a segment contained in K. Assume the property is true any $n \leq N$. For $n = N + 1$, we have two cases to consider. Case one, K has no interior points. Then as argued in the proof of Theorem 10, K is contained in a subspace of X with dimension less than n. By induction, K is closed in the topological sense and hence $x \in K$. Case two, K has at least one interior point x_0. In this case, all the points on the open segment (x_0, x) must be in K. Indeed, assume not, then there exists an $x^* \in (x_0, x)$ such that the open segment $(x_0, x^*) \subseteq K$, but $(x^*, x] \cap K = \emptyset$. Since x_0 is an interior point of K, we can find n linearly independent vectors $e_1, \cdots, e_n$ so that $x_0 + e_i \in K$, $i = 1, \cdots, n$. For any x_k sufficiently close to x, the cone with x_k as the vertex and $x_0 + e_1, \cdots, x_0 + e_n$ as the base necessarily intersects with $(x^*, x]$. So such an $x^* \in (x_0, x)$ with $(x^*, x] \cap K = \emptyset$ does not exist. Therefore $(x_0, x) \subseteq K$ and by definition of being closed in the linear sense, we conclude $x \in K$. □

(3) If K is bounded in the linear sense, it is bounded in the topological sense.

Proof. Assume K is not bounded in the topological sense, then we can find a sequence $(x_k)_{k=1}^\infty$ such that $\|x_k\| \to \infty$. We shall show K is not bounded in the linear sense. Indeed, if the dimension $n = 1$, this is clearly true. Assume the claim is true for any $n \leq N$. For $n = N + 1$, we have two cases to consider. Case one, K has no interior points. Then as argued in the proof of Theorem 10, K is contained in a subspace of X with dimension less than n. By induction, K is not bounded in the linear sense. Case two, K has at least one interior point x_0. Denote by y_k the intersection of the segment $[x_0, x_k]$ with the sphere $S(x_0, 1) = \{z : \|z - x_0\| = 1\}$. For k large enough, y_k always exists. Since a sphere in finite-dimensional space is compact, we can assume without loss of generality that $y_k \to y \in S(x_0, 1)$. Then by an argument similar to that of part (2) (the argument based on cone), the ray starting with x_0 and going through y is contained in K. So K is not bounded in the linear sense. □

13 The Duality Theorem

1.

Proof. Let $Y = \{y : y = \sum_{i=1}^m p_jy_j, p_j \geq 0\}$. If $y, y' \in Y$, then

$$ty + (1-t)y' = \sum_{i=1}^m \left[tp_j + (1-t)p_j'\right]y_j \in Y.$$

So Y is a convex set. □

2.

Proof. $\xi x - \xi z = \xi(x - z) \geq 0$. □

3.

Proof. In the proof of Theorem 3, we already showed that there is an admissible p^* for which $\gamma p^* \geq s$ (formula (21)). Since $S \geq \gamma p^* \geq s \geq S$ by formula (16) and (20), the sup in Theorem 3 is obtained at p^*, hence a maximum. To see the inf in Theorem 3 is a minimum, note under the condition that there are admissible p and ξ, Theorem 3 can be written as

$$\sup\{\gamma p : y \geq Yp, p \geq 0\} = \inf\{\xi y : \gamma \leq \xi Y, \xi \geq 0\} \neq \infty.$$

This is equivalent to

$$\inf\{(-\gamma)p : (-y) \leq (-Y)p, p \geq 0\} = \sup\{\xi(-y) : (-\gamma) \geq \xi(-Y), \xi \geq 0\}.$$

By previous argument for $S = \sup_\gamma \gamma p$, we can find ξ^* such that $\xi^* \geq 0$, $(-\gamma) \geq \xi^*(-Y)$ and $\xi^*(-y) = \sup\{\xi(-y) : (-\gamma) \geq \xi(-Y), \xi \geq 0\}$, i.e. $\xi^* \geq 0$, $\gamma \leq \xi^* Y$, and $\xi^* y = \inf\{\xi y : \gamma \leq \xi Y, \xi \geq 0\}$. That is, the inf in Theorem 3 is obtained at ξ^*, hence a minimum. □

14 Normed Linear Spaces

1.

Proof. Trivial and proof is omitted. □

2.

Proof. $|x - z| = |(x - y) + (y - z)| \leq |x - y| + |y - z|$. □

3.

Proof. (i) $|x|_1 \geq 0$, and $|x|_1 = 0$ if and only if each $a_j = 0$, i.e. $x = 0$.
(ii) $|x + y|_1 = \sum |x_j + y_j| \leq \sum |x_j| + \sum |y_j| = |x|_1 + |y|_1$.
(iii) $|kx|_1 = \sum |kx_j| = \sum |k||x_j| = |k||x|_1$. □

4.

Proof. $f(x) = -\ln x$ is a strictly convex function on $(0, \infty)$. So for any $a, b > 0$ with $a \neq b$, we have

$$f(\theta a + (1-\theta)b) \leq \theta f(a) + (1-\theta) f(b),\ \forall \theta \in [0,1],$$

where "=" holds if and only if $\theta a + (1-\theta)b = a$ or b. That is, one of the following three cases occurs: 1) $\theta = 0$; 2) $\theta = 1$; 3) $a = b$.

We note inequality $f(\theta a + (1-\theta)b) \leq \theta f(a) + (1-\theta) f(b)$ is equivalent to $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$, and by letting $a_i = \frac{|x_i|^p}{|x|_p^p}$ and $b_i = \frac{|y_i|^q}{|y|_q^q}$, we have ($\theta = \frac{1}{p}$)

$$\frac{|x_i y_i|}{|x|_p |y|_q} \leq \frac{1}{p}\frac{|x_i|^p}{|x|_p^p} + \frac{1}{q}\frac{|x_i|^q}{|x|_q^q}.$$

Taking summation gives $\sum_i |x_i y_i| \leq |x|_p |y|_q$.

We consider when $\sum_i |x_i y_i| = |x|_p |y|_q$. Since p, q are real positive numbers and since $\frac{1}{p} + \frac{1}{q} = 1$, we must have $p, q \in (0, 1)$. So among the three cases aforementioned, "=" holds in $\sum_i |x_i y_i| \leq |x|_p |y|_q$ if and only if for each i, $\frac{|x_i|^p}{|x|_p^p} = \frac{|y_i|^q}{|y|_q^q}$, that is, $(|x_1|^p, \cdots, |x_n|^p)$ are proportional to $(|y_1|^q, \cdots, |y_n|^q)$.

For $\sum_i x_i y_i = \sum |x_i y_i|$ to hold, we need $x_i y_i = |x_i y_i|$ for each i. This is the same as $\text{sign}(x_i) = \text{sign}(y_i)$ for each i. In summary, we conclude $xy \leq |x|_p |y|_q$ and the "=" holds if and only if $(|x_1|^p, \cdots, |x_n|^p)$ and $(|y_1|^q, \cdots, |y_n|^q)$ are proportional to each other and $\text{sign}(x_i) = \text{sign}(y_i)$ $(i = 1, \cdots, n)$. □

5.

Proof. Given x, we note

$$|x|_p = |x|_\infty \left(\sum_{i=1}^{n} \frac{|x_i|^p}{|x|_\infty^p} \right)^{1/p} \text{ and } 1 \leq \left(\sum_{i=1}^{n} \frac{|x_i|^p}{|x|_\infty^p} \right)^{1/p} \leq n^{1/p}.$$

Letting $p \to \infty$, we can see $|x|_p \to |x|_\infty$. □

6.

Proof. Every linear subspace of a finite-dimensional normed linear space is again a finite-dimensional normed linear space. So the problem is reduced to proving any finite-dimensional normed space is closed. Fix a basis $e_1, \cdots, e_n$, we introduce the following norm: if $x = \sum a_j e_j$, $\|x\| := (\sum_j a_j^2)^{1/2}$. Then the original norm $|\cdot|$ is equivalent to $\|\cdot\|$. So $(x_k)_{k=1}^\infty$ is a Cauchy sequence under $|\cdot|$ if and only if $\{(a_{k1}, \cdots, a_{kn})\}_{k=1}^\infty$ is a Cauchy sequence in $\mathbb{C}^n$ or $\mathbb{R}^n$. Here $x_k = \sum_{j=1}^n a_{kj} e_j$. Since $\mathbb{C}^n$ and $\mathbb{R}^n$ are complete, we conclude there exists $(b_1, \cdots, b_n) \in \mathbb{C}^n$ or $\mathbb{R}^n$, so that $x_k \to x = \sum b_j e_j$ in $\|\cdot\|$ and hence in $|\cdot|$. □

7.

Proof. If $(z_k)_{k=1}^\infty \subset Y$ is such that $|x - z_k| \to d := \inf_{y \in Y} |x - y|$, then for k sufficiently large, $|z_k| \leq |z_k - x| + |x| \leq 2d + |x|$. Note that $Y = \text{span}\{y_1, \cdots, y_n\}$ is a finite dimensional space, by Theorem 3 (ii), $(z_k)_{k=1}^\infty$ has a subsequence which converges to a point $y_0 \in Y$. Then $\inf_{y \in Y} |x - y|$ is obtained at y_0. □

8.

Proof. (i) Positivity: $|\xi|' = 0$ implies $\xi x = 0$, $\forall x$ with $|x| = 1$. So for any y with $y \neq 0$, $\xi y = |y| \xi (y/|y|) = 0$, i.e. $\xi \equiv 0$. So $|\xi|' = 0$ implies $\xi = 0$, which is equivalent to $\xi \neq 0$ implies $|\xi|' > 0$. $|0| = 0$ is obvious.

(ii) Subadditivity: $|\xi_1 + \xi_2| = \sum_{|x|=1} (\xi_1 + \xi_2) x \leq \sup_{|x|=1} \xi_1 x + \sup_{|x|=1} \xi_2 x = |\xi_1| + |\xi_2|$.

(iii) Homogeneity: $|k\xi| = \sup_{|x|=1} k\xi x = |k| \sup_{|x|=1} \xi x = |k||\xi|$. □

9. (i)

Proof. By formula (47) and (48), it suffices to prove the equality for positive rational r. Suppose $r = \frac{q}{p}$ with $p, q \in \mathbb{Z}^+$. By formula (49) and by induction, we have

$$\overbrace{(x, y) + \cdots + (x, y)}^{n} = (nx, y)\,.$$

Therefore

$$p(rx, y) = (prx, y) = (qx, y) = q(x, y), \text{ i.e. } (rx, y) = \frac{q}{p}(x, y) = r(x, y).$$

□

(ii)

Proof. For any given k, we can find a sequence of rational numbers $(r_n)_{n=1}^\infty$ such that $r_n \to k$ as $n \to \infty$. Then $k(x, y) = \lim_{n\to\infty} r_n(x, y) = \lim_{n\to\infty}(r_n x, y) = (\lim_{n\to\infty} r_n x, y) = (kx, y)$, where the third "=" uses the fact that $(\cdot, y)$ defines a continuous linear functional on X. □

15 Linear Mappings Between Normed Linear Spaces

1.

Proof. By Lemma 1, $|Tx_n - Tx| = |T(x_n - x)| \le c|x_n - x|$. So if $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} Tx_n = Tx$. □

2.

Proof. Suppose $|T_n - T|$ does not tend to zero. Then there exists $\varepsilon > 0$ and a sequence $(x_n)_{n=1}^\infty$ such that $|x_n| = 1$ and $|(T_n - T)x_n| \ge \varepsilon$. By Theorem 3(ii), we can without loss of generality assume $(x_n)_{n=1}^\infty$ converges to some point x^*. Then

$$|(T_n - T)x_n| - |(T_n - T)x^*| \le |(T_n - T)x_n - (T_n - T)x^*| \le |T(x_n - x^*)| + |T_n(x_n - x^*)| \le |T||x_n - x^*| + |T_n||x_n - x^*|.$$

For n sufficiently large, $|(T_n - T)x_n| - |(T_n - T)x^*|$ will be greater than $\varepsilon/2$, while $|T||x_n - x^*| + |T_n||x_n - x^*|$ will be as small as we want, provided that we can prove $(|T_n|)_{n=1}^\infty$ is bounded. Indeed, this is the *principle of uniform boundedness* (see, for example, Lax [6], Chapter 10, Theorem 3). Thus we have arrived at a contradiction which shows our assumption is wrong. □

Remark 14. *Can we find an elementary proof without using the principle of uniform boundedness in functional analysis, especially since we are working with finite dimensional space?*

3.

Proof. First of all, $\sum_{k=0}^\infty R^k$ is well-defined, since by $|R| < 1$, $\left(\sum_{k=0}^K R^k\right)_{K=0}^\infty$ is a Cauchy sequence in X'. Then we note $S\sum_{k=0}^\infty R^k = \sum_{k=0}^\infty R^k - \sum_{k=1}^\infty R^k = I$ and $(\sum_{k=0}^\infty R^k)S = \sum_{k=0}^\infty R^k - \sum_{k=1}^\infty R^k = I$. So S is invertible and $S^{-1} = \sum_{k=0}^\infty R^k$. □

4.

Proof. Assume all the conditions in Theorem 5. Define $R = -T^{-1}(S - T)$, then $|R| \le |T^{-1}||S - T| < 1$. So by Theorem 6, $I - R = T^{-1}S$ is invertible, hence $S = T \circ T^{-1}S$ is invertible. □

5.

Proof. If for some m, $|R^m| < 1$, we define $U = \sum_{k=0}^\infty R^{km} = I + R^m U$. U is well-defined, and the following linear map is also well-defined: $V = U + RU + \cdots + R^{m-1}U$. Then $SV = U + RU + \cdots + R^{m-1}U - (RU + R^2U + \cdots + R^m U) = U - R^m U = I$. This shows S is invertible. □

6.

Proof. For any $x \in \mathbb{R}^n$, $|Tx|_\infty = \max_i |\sum_{j=1}^n t_{ij}x_j| \le \max_i(\sum_{j=1}^n |t_{ij}|)|x|_\infty$. So $|T| = \sup_{x\ne 0}\frac{|Tx|_\infty}{|x|_\infty} \le \max_i \sum_j |t_{ij}|$. For the other direction, suppose $\sum_j |t_{i_0 j}| = \max_i \sum_j |t_{ij}|$ and we choose

$$x^* = (\mathrm{sign}(t_{i_0 1}), \cdots, \mathrm{sign}(t_{i_0 n}))^T,$$

then $|x^*|_\infty = 1$ and $Tx^* = (\sum_j t_{1j}x_j^*, \cdots, \sum_j t_{i_0 j}x_j^*, \cdots, \sum_j t_{nj}x_j^*)^T$. So $|Tx^*|_\infty \ge \sum_j |t_{i_0 j}| = \max_i \sum_j |t_{ij}|$. This implies $|T| \ge \max_i \sum_j |t_{ij}|$. Combined, we conclude $|T| = \max_i \sum_j |t_{ij}|$. □

7.

Proof. For any $x = (x_1, \cdots, x_n)'$, we have

$$|Tx|_l = \sum_{i=1}^n \left|\sum_{j=1}^n t_{ij}x_j\right| \le \sum_{i=1}^n\sum_{j=1}^n |t_{ij}x_j| \le \sum_{i,j} |t_{ij}||x|_\infty.$$

So $|T| = \sup_{|x|_\infty = 1}\frac{|Tx|_l}{|x|_\infty} \le \sum_{i,j}|t_{ij}|$. □

8.

Proof. The proof is very similar to the content on page 97, Chapter 7, the material up to Theorem 18. So we omit the solution. □

16 Positive Matrices

1.

Proof. Let $x^* = (1, \cdots, 1)^T$ and $\lambda^* = \max_{1\le i\le n}\sum_{j=1}^n p_{ij}$, then $Px^* \le \lambda^* x^*$. So $t(P) \ne \emptyset$ and $t^*(P) = \{0 \le \lambda \le \lambda^* : \lambda \in t(P)\}$ is a bounded, nonempty set. We show further $t^*(P)$ is closed. Suppose $(\lambda^m)_{m=1}^\infty \subset t^*(P)$ converges to a point λ. Denote by x^m the nonnegative and nonzero vector such that $Px^m \le \lambda^m x^m$. Without loss of generality, we assume $\sum_{i=1}^n x_i^m = 1$. Then $(x^m)_{m=1}^\infty$ is bounded and we can assume $x^m \to x$ for some $x \ge 0$ with $\sum_{i=1}^n x_i = 1$. Passing to limit gives us $Px \le \lambda x$. Clearly $0 \le \lambda \le \lambda^*$. So $\lambda \in t^*(P)$. This shows $t^*(P)$ is compact and $t(P)$ has a minimum $\bar\lambda$.

Denote by $\bar x$ the nonzero and nonnegative vector such that $P\bar x \le \bar\lambda \bar x$. We show we actually have $P\bar x = \bar\lambda \bar x$. Assume not, there must exist some $k \in \{1, \cdots, n\}$ such that $\sum_{j=1}^n p_{ij}\bar x_{ij} \le \bar\lambda \bar x_i$ for $i \ne k$ and $\sum_{j=1}^n p_{kj}\bar x_j < \lambda x \lambda x_k$. Consider the vector $\widehat{x} = \bar x - \varepsilon e_k$, where $\varepsilon > 0$ and e_k has the k-th component equal to 1 with all the other components zero. Then in the inequality $P\bar x \le \bar\lambda \bar x$, each component of LHS is decreased when $\bar x$ is replaced by $\widehat{x}$, while only the k-th component of RHS is decreased by an amount of $\lambda\varepsilon$. So for ε small enough, $P\widehat{x} < \bar\lambda\widehat{x}$, and we can find a $\widehat{\lambda} < \bar\lambda$ such that $P\widehat{x} \le \widehat{\lambda}\widehat{x}$. Note $\bar\lambda > 0$ (otherwise $\bar x = 0$, a contradiction), so we can also let $\widehat{\lambda} > 0$. This contradicts with $\bar\lambda = \min_{\lambda\in t(P)}\lambda$. We have shown $\bar\lambda > 0$ is an eigenvalue of P which has a nonzero, nonnegative eigenvalue. By Theorem 1(iv), $\bar\lambda = \lambda(P)$. □

2.

Proof. P^m has a dominant positive eigenvalue λ_0. By Spectral Mapping Theorem, there is an eigenvalue λ of P, such that $\lambda^m = \lambda_0$. Suppose λ is real, then we can further assume $\lambda > 0$ by replacing λ with $-\lambda$ if necessary. Then for any other eigenvalue λ' of P, Spectral Mapping Theorem implies $(\lambda')^m$ is an eigenvalue of P^m. So $|(\lambda')^m| < \lambda_0 = \lambda^m$, i.e. $|\lambda'| < \lambda$.

To show we can take λ as real, denote by x the eigenvector of P^m associated with λ_0. Then

$$P^m x = \lambda_0 x.$$

Let P act on this relation:

$$P^{m+1}x = P^m(Px) = \lambda_0 Px.$$

This shows Px is too an eigenvector of P^m with eigenvalue λ_0. By Theorem 1(iv), $Px = cx$ for some positive number c. Repeated application of P shows that $P^m x = c^m x$. Therefore $c^m = \lambda_0$. Let $\lambda = c$. □

17 How to Solve Systems of Linear Equations

The three-term recursion formulae

$$x_{n+1} = (s_n A + p_n I)x_n + q_n x_{n-1} - s_n b$$

is introduced by Rutishauser et al. [1]. See Papadrakakis [9] for a survey on a family of vector iterative methods with three-term recursion formulae and Golub and van Loan [2] for a gentle introduction to the Chebyshev semi-iterative method (section §10.1.5).

1.

Proof. $I = AA^{-1}$. So $1 = |I| = |AA^{-1}| \le |A||A^{-1}| = \kappa(A)$. □

2.

Proof. If we use the method of Section 1, we need to solve for N the following inequality: $\frac{2}{\alpha}(1-\frac{1}{\kappa})^N F(x_0) < 10^{-3}$. Plug in numbers, we have $N > 757$. If we use the method of Section 2, we need to solve for N the inequality $2(1+\frac{2}{\sqrt{\kappa}})^{-N}\|e_0\| < 10^{-3}$. Plug in numbers, we have $N > 42$. The numbers of steps needed in respective methods differ a great deal. □

3.

Proof. We first summarize the algorithm. We need to solve the system of linear equations $Ax = b$, where b is a given vector and A is an invertible matrix. We start with an initial guess x_0 of the solution and define $r_0 = Ax_0 - b$. TO BE CONTINUED ... $\square$

4. We solve the following problem from the first edition of the textbook: Use the computer program in Exercise 3 to solve the system of equations

$$Ax = f, \ A_{ij} = c + \frac{1}{i+j+1}, \ f_i = \frac{1}{i!},$$

c some nonnegative constant. Vary c between 0 and 1, and the order K of the system between 5 and 20.

Proof. TO BE CONTINUED ... $\square$

18 How to Calculate the Eigenvalues of Self-Adjoint Matrices

TO BE CONTINUED ...

19 Appendix

19.1 Special Determinants

1.

Proof. The system of equations $p(a_i) = p_i$ $(i = 1, \cdots, n)$ can be written as

$$\begin{pmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ 1 & a_2 & \cdots & a_2^{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & a_n & \cdots & a_n^{n-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ p_n \end{pmatrix}$$

Since $a_1, \cdots, a_n$ are distinct, by the formula for the determinant of Vandermonde matrix, the determinant of the matrix is equal to $\prod_{j>i}(a_j - a_i)$. $\square$

2.

Proof. Denote the matrix by A. We claim $\det A = \frac{\prod_{j>i}(a_j-a_i)^2}{\prod_{i,j}(1+a_ia_j)}$. Indeed, by subtracting column 1 from each of the other columns, we have

$$\begin{aligned}
\det A &= \det \begin{bmatrix} \frac{1}{1+a_1^2} & \frac{1}{1+a_1a_2} & \frac{1}{1+a_1a_3} & \cdots & \frac{1}{1+a_1a_n} \\ \frac{1}{1+a_2a_1} & \frac{1}{1+a_2^2} & \frac{1}{1+a_2a_3} & \cdots & \frac{1}{1+a_2a_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{1+a_ia_1} & \frac{1}{1+a_ia_2} & \frac{1}{1+a_ia_3} & \cdots & \frac{1}{1+a_ia_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{1+a_na_1} & \frac{1}{1+a_na_2} & \frac{1}{1+a_na_3} & \cdots & \frac{1}{1+a_n^2} \end{bmatrix} \\
&= \det \begin{bmatrix} \frac{1}{1+a_1^2} & \frac{a_1(a_2-a_1)}{(1+a_1^2)(1+a_1a_2)} & \frac{a_1(a_3-a_1)}{(1+a_1^2)(1+a_1a_3)} & \cdots & \frac{a_1(a_n-a_1)}{(1+a_1^2)(1+a_1a_n)} \\ \frac{1}{1+a_2a_1} & \frac{a_2(a_2-a_1)}{(1+a_2a_1)(1+a_2^2)} & \frac{a_2(a_3-a_1)}{(1+a_2a_1)(1+a_2a_3)} & \cdots & \frac{a_2(a_n-a_1)}{(1+a_2a_1)(1+a_2a_n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{1+a_ia_1} & \frac{a_i(a_2-a_1)}{(1+a_ia_1)(1+a_ia_2)} & \frac{a_i(a_3-a_1)}{(1+a_ia_1)(1+a_ia_3)} & \cdots & \frac{a_i(a_n-a_1)}{(1+a_ia_1)(1+a_ia_n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{1+a_na_1} & \frac{a_n(a_2-a_1)}{(1+a_na_1)(1+a_na_2)} & \frac{a_n(a_3-a_1)}{(1+a_na_1)(1+a_na_3)} & \cdots & \frac{a_n(a_n-a_1)}{(1+a_na_1)(1+a_n^2)} \end{bmatrix}
\end{aligned}$$

By extracting the common factor $\frac{1}{1+a_i a_i}$ $(i = 1, \cdots, n)$ from each row and $(a_j - a_1)$ $(j = 2, \cdots, n)$ from each column, we have

$$\det A = \frac{\prod_{j=2}^n (a_j - a_1)}{\prod_{i=1}^n (1 + a_i a_1)} \det \begin{bmatrix} 1 & \frac{a_1}{1+a_1a_2} & \frac{a_1}{1+a_1a_3} & \cdots & \frac{a_1}{1+a_1a_n} \\ 1 & \frac{a_2}{1+a_2^2} & \frac{a_2}{1+a_2a_3} & \cdots & \frac{a_2}{1+a_2a_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \frac{a_i}{1+a_ia_2} & \frac{a_i}{1+a_ia_3} & \cdots & \frac{a_i}{1+a_ia_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \frac{a_n}{1+a_na_2} & \frac{a_n}{1+a_na_3} & \cdots & \frac{a_n}{1+a_n^2} \end{bmatrix}$$

Subtracting row 1 from each of the other rows, we get

$$\det A = \frac{\prod_{j=2}^n (a_j - a_1)}{\prod_{i=1}^n (1 + a_i a_1)} \det \begin{bmatrix} 1 & \frac{a_1}{1+a_1a_2} & \frac{a_1}{1+a_1a_3} & \cdots & \frac{a_1}{1+a_1a_n} \\ 0 & \frac{a_2-a_1}{(1+a_1a_2)(1+a_2^2)} & \frac{a_2-a_1}{(1+a_1a_3)(1+a_2a_3)} & \cdots & \frac{a_2-a_1}{(1+a_1a_n)(1+a_2a_n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \frac{a_i-a_1}{(1+a_1a_2)(1+a_ia_2)} & \frac{a_i-a_1}{(1+a_1a_3)(1+a_ia_3)} & \cdots & \frac{a_i-a_1}{(1+a_1a_n)(1+a_ia_n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \frac{a_n-a_1}{(1+a_1a_2)(1+a_na_2)} & \frac{a_n-a_1}{(1+a_1a_3)(1+a_na_3)} & \cdots & \frac{a_n-a_1}{(1+a_1a_n)(1+a_n^2)} \end{bmatrix}$$

By the Laplace expansion and extracting the common factor $(a_i - a_1)$ from row 2 through n and the common factor $\frac{1}{1+a_1a_j}$ from column 2 through n, we get

$$\det A = \frac{\prod_{j=2}^n (a_j - a_1)^2}{\prod_{i=1}^n (1 + a_i a_1) \prod_{j=2}^n (1 + a_1 a_j)} \det \begin{bmatrix} \frac{1}{1+a_2^2} & \frac{1}{1+a_2a_3} & \cdots & \frac{1}{1+a_2a_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{1+a_ia_2} & \frac{1}{1+a_ia_3} & \cdots & \frac{1}{1+a_ia_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{1+a_na_2} & \frac{1}{1+a_na_3} & \cdots & \frac{1}{1+a_n^2} \end{bmatrix}$$

By induction, we can prove our claim. □

19.2 The Pfaffian

1.

Proof. By the Laplace expansion and Exercise 16 of Chapter 5, we have

$$\begin{aligned} \det \begin{bmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{bmatrix} &= a \det \begin{bmatrix} a & b & c \\ -d & 0 & f \\ -e & -f & 0 \end{bmatrix} - b \det \begin{bmatrix} a & b & c \\ 0 & d & e \\ -e & -f & 0 \end{bmatrix} + c \det \begin{bmatrix} a & b & c \\ 0 & d & e \\ -d & 0 & f \end{bmatrix} \\ &= a(-bef + cdf + af^2) - b(-be^2 + cde + aef) + c(adf - bde + cd^2) \\ &= -2abef + 2acdf - 2bcde + a^2f^2 + b^2e^2 + c^2d^2 \\ &= (af - be + cd)^2. \end{aligned}$$

□

19.3 Symplectic Matrices

1.

Proof. We work by induction. For $n = 1$, A has the form of $\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}$. Since $\det A \neq 0$, $a \neq 0$. We note

$$\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{a} \end{bmatrix} \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{a} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Now assume the claim is true for $1, \cdots, n$, we show it also holds for $n+1$. Indeed, we write A into the form

$$\begin{bmatrix} 0 & a & * \\ -a & 0 & * \\ * & * & A_1 \end{bmatrix},$$

where A_1 is a $2n \times 2n$ anti-self-adjoint matrix. Then

$$\begin{bmatrix} 1 & 0 & 0_{1\times 2n} \\ 0 & \frac{1}{a} & 0_{1\times 2n} \\ 0_{2n\times 1} & 0_{2n\times 1} & I_{2n\times 2n} \end{bmatrix} \begin{bmatrix} 0 & a & * \\ -a & 0 & * \\ * & * & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0_{1\times 2n} \\ 0 & \frac{1}{a} & 0_{1\times 2n} \\ 0_{2n\times 1} & 0_{2n\times 1} & I_{2n\times 2n} \end{bmatrix} = \begin{bmatrix} 0 & 1 & * \\ -1 & 0 & * \\ * & * & A_1 \end{bmatrix}.$$

Recall that multiplying an elementary matrix from the left is equivalent to an elementary row manipulation, while multiplying an elementary matrix from the right is equivalent to an elementary column manipulation. A being anti-self-adjoint implies $A_{ij} = -A_{ji}$, so we can find a sequence of elementary matrices $U_1, U_2, \cdots, U_k$ such that

$$U_k \cdots U_2 U_1 \begin{bmatrix} 0 & 1 & * \\ -1 & 0 & * \\ * & * & A_1 \end{bmatrix} U_1^T U_2^T \cdots U_k^T = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & A_1 \end{bmatrix}.$$

By assumption, $A_1 = F_1 J_1 F_1^T$ for some real matrix F_1 with $\det F_1 \neq 0$ and $J_1 = \begin{bmatrix} 0_{n\times n} & I_{n\times n} \\ -I_{n\times n} & 0_{n\times n} \end{bmatrix}$. Therefore $(U := U_k \cdots U_2 U_1)$

$$\begin{bmatrix} I_{2\times 2} & 0 \\ 0 & F_1^{-1} \end{bmatrix} U \begin{bmatrix} 0 & 1 & * \\ -1 & 0 & * \\ * & * & A_1 \end{bmatrix} U^T \begin{bmatrix} I_{2\times 2} & 0 \\ 0 & (F_1^{-1})^T \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & J_1 \end{bmatrix}.$$

Define

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & I_{n\times n} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & I_{n\times n} & 0 \end{bmatrix}$$

and

$$F = \left(L \begin{bmatrix} I_{2\times 2} & 0 \\ 0 & F_1^{-1} \end{bmatrix} U \begin{bmatrix} 1 & 0 & 0_{1\times 2n} \\ 0 & \frac{1}{a} & 0_{1\times 2n} \\ 0_{2n\times 1} & 0_{2n\times 1} & I_{2n\times 2n} \end{bmatrix} \right)^{-1}.$$

Then

$$\begin{aligned} F^{-1} A (F^{-1})^T &= L \begin{bmatrix} I_{2\times 2} & 0 \\ 0 & F_1^{-1} \end{bmatrix} U \begin{bmatrix} 1 & 0 & 0_{1\times 2n} \\ 0 & \frac{1}{a} & 0_{1\times 2n} \\ 0_{2n\times 1} & 0_{2n\times 1} & I_{2n\times 2n} \end{bmatrix} \begin{bmatrix} 0 & a & * \\ -a & 0 & * \\ * & * & A_1 \end{bmatrix} \\ &\quad \cdot \begin{bmatrix} 1 & 0 & 0_{1\times 2n} \\ 0 & \frac{1}{a} & 0_{1\times 2n} \\ 0_{2n\times 1} & 0_{2n\times 1} & I_{2n\times 2n} \end{bmatrix} U^T \begin{bmatrix} I_{2\times 2} & 0 \\ 0 & (F_1^{-1})^T \end{bmatrix} L^T \\ &= \begin{bmatrix} 0 & I_{(n+1)\times(n+1)} \\ -I_{(n+1)\times(n+1)} & 0 \end{bmatrix}. \end{aligned}$$

By induction, we have proved the claim. □

2.

Proof. For any given x and y, define $f(t) = (S(t)x, JS(t)y)$. Then we have

$$\begin{aligned}
\frac{d}{dt}f(t) &= (\frac{d}{dt}S(t)x, JS(t)y) + (S(t)x, J\frac{d}{dt}S(t)y) \\
&= (G(t)S(t)x, JS(t)y) + (S(t)x, JG(t)S(t)y) \\
&= (JL(t)S(t)x, JS(t)y) + (S(t)x, J^2L(t)S(t)y) \\
&= (L(t)S(t)x, J^TJS(t)y) - (S(t)x, L(t)S(t)y) \\
&= (S(t)x, L(t)S(t)y) - (S(t)x, L(t)S(t)y) \\
&= 0.
\end{aligned}$$

So $f(t) = f(0) = (S(0)x, JS(0)y) = (x, Jy)$. Since x and y are arbitrary, we conclude $S(t)$ is a family of symplectic matrices. □

4.

Proof. We note

$$\frac{dv}{dt} = \begin{bmatrix} \sum_{i=1}^{2n} \frac{\partial v_1}{\partial u_i}\frac{du_i}{dt} \\ \cdots \\ \sum_{i=1}^{2n} \frac{\partial v_{2n}}{\partial u_i}\frac{du_i}{dt} \end{bmatrix} = \frac{\partial v}{\partial u}\frac{du}{dt} = \frac{\partial v}{\partial u}JH_u.$$

$H(u)$ can be seen as a function of v: $K(v) \triangleq H(u(v))$. So

$$K_v = \begin{bmatrix} \frac{\partial K}{\partial v_1} \\ \cdots \\ \frac{\partial K}{\partial v_{2n}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{2n} \frac{\partial H}{\partial u_i}\frac{\partial u_i}{\partial v_1} \\ \cdots \\ \sum_{i=1}^{2n} \frac{\partial H}{\partial u_i}\frac{\partial u_i}{\partial v_{2n}} \end{bmatrix} = \begin{bmatrix} \frac{\partial u_1}{\partial v_1} & \cdots & \frac{\partial u_{2n}}{\partial v_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial u_1}{\partial v_{2n}} & \cdots & \frac{\partial u_{2n}}{\partial v_{2n}} \end{bmatrix} \begin{bmatrix} \frac{\partial H}{\partial u_1} \\ \cdots \\ \frac{\partial H}{\partial u_{2n}} \end{bmatrix} = \left(\frac{\partial u}{\partial v}\right)^T H_u.$$

Since $\partial v/\partial u$ is symplectic, by Theorem 2, $\partial u/\partial v$ and $(\partial v/\partial u)^T$ are also symplectic. So using formula (4) gives us

$$\frac{dv}{dt} = \frac{\partial v}{\partial u}J\left(\frac{\partial v}{\partial u}\right)^T\left(\frac{\partial u}{\partial v}\right)^T H_u = J\left(\frac{\partial u}{\partial v}\right)^T H_u = JK_v.$$

□

19.4 Tensor Product

19.5 Lattices

19.6 Fast Matrix Multiplication

There are no exercise problems for this section. For examples of implementation of Strassen's algorithm, we refer to Huss-Lederman et al. [7] and references therein.

19.7 Gershgorin's Theorem

1.

Proof. Using the notation of Gershgorin Circle Theorem, let $B(t) = D + tF$, $t \in [0, 1]$. The eigenvalues of $B(t)$ are continuous functions of t (Theorem 6 of Chapter 9). For $t = 0$, the eigenvalues of $B(0)$ are the diagonal entries of A. As t goes from 0 to 1, the radius of Gershgorin circles corresponding to $B(t)$ become bigger while the centers remain the same. So we can find for each d_i a continuous path $\gamma_i(t)$ such that $\gamma_i(0) = d_i$ and $\gamma_i(t)$ is an eigenvalue of $B(t)$ ($0 \le t \le 1$, $i = 1, \cdots, n$). Moreover, by Gershgorin Circle Theorem, each path $\gamma_i(t)$ ($0 \le t \le 1$) is contained in disc $C_i = \{x : |x - d_i| \le |f_i|_l\}$. If for some i_1 and i_2 with $i_1 \ne i_2$, $\gamma_{i_2}(0)$ falls into C_{i_1}, then it's necessary that $C_{i_1} \cap C_{i_2} \ne \emptyset$. This implies that for any Gershgorin disc that is disjoint from all the other Gershgorin discs, there is one and only one eigenvalue of A falls within it.

Remark 15. *There's a strengthened version of Gershgorin Circle Theorem that can be found at Wikipedia (http://en.wikipedia.org/wiki/Gershgorin_circle_theorem). The above exercise problem's solution is an adaption of the proof therein. The claim: If the union of k Gershgorin discs is disjoint from the union of the other $(n-k)$ Gershgorin discs, then the former union contains exactly k and the latter $(n-k)$ eigenvalues of A.*

□

19.8 The Multiplicity of Eigenvalues

19.9 The Fast Fourier Transform

There are no exercise problems for this section.

19.10 The Spectral Radius

Remark 16. *In the textbook (pp. 340), when the author applied Cauchy integral theorem to get formula (20): $\int_{|z|=s} R(z)z^j dz = 2\pi i A^j$, he used the version of Cauchy integral theorem for the outside region of a simple closed curve (see, for example, Gong and Gong [4], Chapter 2, Exercise 8).*

1.

Proof. If T is an upper triangular matrix with diagonal entries $a_1, \cdots, a_n$, then its characteristic polynomial $p_T(\lambda) = \det(\lambda I - T) = \prod_{i=1}^n (\lambda - a_i)$. So

$$\lambda_0 \text{ is an eigenvalue of } T \Leftrightarrow \det(\lambda_0 I - T) = 0 \Leftrightarrow \prod_{i=1}^n (\lambda_0 - a_i) = 0 \Leftrightarrow \lambda_0 \text{ is equal to one of } a_1, \cdots, a_n.$$

□

2.

Proof. Let $D = \text{diag}\{a_1, \cdots, a_n\}$. Then $De_i = \text{diag}\{0, \cdots, 0, a_i, 0, \cdots, 0\}$. So $\|De_i\| = |a_i|$. This shows $\|D\| \ge \max_{1\le i\le n} |a_i|$. For any $x \in \mathbb{C}^n$, $Dx = \text{diag}\{a_1x_1, \cdots, a_nx_n\}$. So $\|Dx\| = \sqrt{\sum_{i=1}^n |a_ix_i|^2} \le \max_{1\le i\le a} |a_i| \cdot \|x\|$. So $\|D\| \le \max_{1\le i\le n} |a_i|$. Combined, we conclude $\|D\| = \max_{1\le i\le n} |a_i|$. □

3.

Proof. By examining the proof for Euclidean space, we see inner product is not really used. All that has been exploited is just norm. So the proof for any finite-dimensional normed linear space is entirely identical to that of finite-dimensional Euclidean space. □

4.

Proof. It suffices to note that a sequence $(A_n)_{n=1}^\infty$ converges to A in matrix norm iff each $((A_n)_{ij})_{n=1}^\infty$ converges to A_{ij} (Exercise 16 of Chapter 7). □

5.

Proof. By formula (16) of Chapter 5: $D(a_1, \cdots, a_n) = \sum \sigma(p) a_{p_1 1} \cdots a_{p_n n}$, we conclude the determinant of any analytic matrix (i.e. matrix-valued analytic function) is analytic. By Cramer's rule and $\det A(z) \neq 0$ in G, we conclude $A^{-1}(z)$ is also analytic in G. □

6.

Proof. By Exercise 4, Cauchy integral theorem for matrix-valued functions is reduced to Cauchy integral theorem for each entry of an analytic matrix. □

19.11 The Lorentz Group

19.12 Compactness of the Unit Ball

1. (i)

Proof. We use the notation of Theorem 3. For simplicity, we assume G is convex so that for any $x, y \in G$, the segment $\{z : z = (1-t)x + ty\} \subset G$. Then by Mean Value Theorem, there exists $c \in (0,1)$ such that

$$|f(x) - f(y)| = |\nabla f((1-c)x + cy) \cdot (y - x)| \leq dm|x - y|,$$

where d is the dimensional of the Euclidean space in which G resides. This shows the elements of D are equi-continuous in G. □

(ii)

Proof. From (i), we know each element of D is uniformly continuous. So they can be extended to $\bar{G}$, the closure of G. Then Theorem 3 is the result of the following version of Arzela-Ascoli Theorem (see, for example, Yosida [12]): *Let S be a compact metric space, and $C(S)$ the Banach space of (real- or) complex-valued continuous functions $x(s)$ normed by $\|x\| = \sup_{s\in S}|x(s)|$. Then a sequence $\{x_n(s)\} \subset C(S)$ is relatively compact in $C(S)$ if the following two conditions are statisfied: (a) $\{x_n(s)\}$ is uniformly bounded; (b) $\{x_n(s)\}$ is equi-continuous.* □

19.13 A Characterization of Commutators

There are no exercise problems for this section.

19.14 Liapunov's Theorem

1.

Proof. This is basically about how to extend the Riemann integral to Banach space valued functions. The theory is essential the same as the scalar case – just replace the Euclidean norm with an arbitrary norm. So we omit the details. □

2.

Proof. It suffices to note $A_n \to A$ in matrix norm if and only if each entry of A_n converges to the corresponding entry of A (see Exercise 7 and formula (51) of Chapter 7). □

3.

Proof. We note for $T' > T$, by Definition 2

$$\left\| \int_T^{T'} e^{W^*t} e^{Wt} dt \right\| \leq \int_T^{T'} \left\| e^{W^*t} \right\| \left\| e^{Wt} \right\| dt.$$

By Lemma 4, $\lim_{T,T'\to\infty} \int_T^{T'} \left\| e^{W^*t} \right\| \left\| e^{Wt} \right\| dt = 0$. So by Cauchy's criterion, we conclude the integral (12) exists. □

19.15 The Jordan Canonical Form

There are no exercise problems for this section. In the below, we give a brief summary of how to compute the Jordan canonical form from the standpoint of module theory.

Oftentimes, we need to compute the Jordan canonical form of a given square matrix A, e.g. in solving system of linear ordinary differential equations. The method given in the textbook, although conceptually appealing, is still complicated in practice. We hope to find an easier method.

Qiu [11] (Reading Materials 3, 5, 7) explained a way to compute the Jordan canonical form of A through the elementary divisors of the λ-matrix $A(\lambda) := \lambda I - A$. This is essentially the factorization theorem of a finitely generated module over the principal ideal domain $K[\lambda]$, where $K[\lambda]$ is the polynomial ring over the field K. A systematic exposition of this theory can be found in Nie and Ding [8] (Chapter 5 and 6) and Gong [3].

In the following, we will summarize the relevant results for future reference. The exposition combines materials from Gong [3] and Qiu [11]: the theoretical aspect is taken from Gong [3], and the computational aspect is taken from Qiu [11].

19.15.1 Introduction

Before we dive into abstract theory, we explain the prototype that motivates the theory. Let V be a finite dimensional linear vector space over the complex field $\mathbb{C}$. Let A be a linear operator on V. We want to choose a suitable set of basis for V such that the matrix representation of A has a nice form, e.g. diagonal, block diagonal, etc.

The key observation is that the matrix of A has a block diagonal form $\text{diag}\{D_1, D_2, \cdots, D_k\}$ if and only if V can be decomposed into the direct sum of subspaces invariant under A:

$$V = V_1 \bigoplus V_2 \bigoplus \cdots \bigoplus V_k.$$

Hamilton-Caley Theorem says $P(A)$ is the zero linear transformation on V (i.e. $P(A)v = 0$, $\forall v \in V$), where $P(\lambda)$ is the characteristic polynomial $|\lambda I - A|$. In other words, V is the kernel of $P(A)$. This immediately reduces the decomposition of V to the factorization of $P(\lambda)$ into irreducible elements in the polynomial ring $K[\lambda]$. Note $K[\lambda]$ is a principal ideal domain, that is, something behaves like the ring of integers. Thus, we will have at our disposal the unique factorization theorem of a principal ideal domain into prime elements.

Extending those concrete objects to an abstract setting, we come to the standpoint of representation theory: *linear vector spaces equipped with linear transformations are modules over principal ideal domains.* Therefore, the (concrete) decomposition of V into direct sum of subspaces invariant under A is generalized to the decomposition of a finitely generated module over a principal ideal domain into the direct sum of primary submodules.

With all these specific objects in mind (linear space, polynomial ring, characteristic polynomial, integers, prime numbers, etc.), we are ready to understand the module theory's standpoint.

19.15.2 Decomposition of a finitely generated module over a principal ideal domain

For the definitions of *ring*, *integral domain* and *principal ideal domain*, we refer to Gong [3], Chapter 1. Roughly speaking, they are abstract generalizations of the integer ring $\mathbb{Z}$ and the polynomial ring $K[\lambda]$ over a field K. Here, the nontrivial conceptual leap is the move from integral domain to principal ideal domain. Integer ring $\mathbb{Z}$ and the polynomial ring $K[\lambda]$ are principal ideal domains because of Eulidean algorithm. The Unique Factorization Theorem for $\mathbb{Z}$ and $K[\lambda]$ can then be generalized to principal ideal domain. That's sufficient for our purpose of decomposing a linear vector space.

The connection between Unique Factorization Theorem of $K[\lambda]$ and the decomposition of V is made by the concept of module. More precisely, define the action of $f(\lambda) \in K[\lambda]$ on V by

$$f(\lambda)v = f(A)v, \ \forall v \in V.$$

If $V = \ker(f(\lambda))$ and $f(\lambda)$ has the factorization into non-associated prime elements $f(\lambda) = \prod_{i=1}^{n} [p_i(\lambda)]^{e_i}$, we have $V = \bigoplus_{i=1}^{n} \ker(p_i^{e_i})$.

For a general module, various technical conditions are needed to produce a clean-cut form of Unique Factorization Theorem. The relevant results are listed below. For all the jargons, results, and proofs, see Gong [3], Chapter 4.

Theorem 1. *(***Cyclic decomposition theorem of a finitely generated module over a principal ideal domain – in terms of elementary divisors***) Suppose M is a non-zero finitely generated module over a principal ideal domain R. Then*

$$M = M_{tor} \bigoplus M_{free},$$

where M_{tor} is the collection of torsion elements in M and M_{free} is a free module, whose rank is uniquely determined by M.

$ann(M) = \{r \in R : rM = \{0\}\}$ is an ideal of R and hence is generated by some element $p_1^{e_1} \cdots p_n^{e_n} \in R$. Here p_i's are prime elements which are not associated to each other. Then

$$M_{tor} = M_{p_1} \bigoplus \cdots \bigoplus M_{p_n},$$

where $M_{p_i} = \{v : p_i^{e_i} v = 0\}$ is a primary module with order $p_i^{e_i}$, $i = 1, \cdots, n$.

Each M_{p_i} can be further decomposed into the sum of cyclic submodules

$$M_{p_i} = C_{i,1} \bigoplus \cdots \bigoplus C_{i,k_i},$$

where $C_{i,j}$ has order $p_i^{e_{i,j}}$ $(j = 1, \cdots, k_i)$ and

$$e_i = e_{i,1} \geq e_{i,2} \geq \cdots \geq e_{i,k_i} \geq 1,\ i = 1, \cdots, n.$$

$p_i^{e_{i,j}}$'s are called **elementary divisors** *and are uniquely determined by M up to the equivalence of associativity. In summary, we have*

$$M = (C_{1,1} \bigoplus \cdots \bigoplus C_{1,k_1}) \bigoplus \cdots \bigoplus (C_{n,1} \bigoplus \cdots \bigoplus C_{n,k_n}) \bigoplus M_{free}.$$

Note if S and T are cyclic submodules of M with $\text{ann}(S) = \langle a \rangle$ and $\text{ann}(T) = \langle b \rangle$, and $\gcd(a, b) = 1$, then $S \cap T = \{0\}$ and $S \bigoplus T$ is also a cyclic submodule with $\text{ann}(S \bigoplus T) = \langle ab \rangle$. With this observation, let $D_1 = C_{1,1} \bigoplus \cdots \bigoplus C_{n,1}$, then D_1 is a cyclic submodule with order $q_1 = \prod_{i=1}^n p_i^{e_{i,1}}$. Define D_2, $\cdots$, D_m similarly, where $m = \max_i(k_i)$. More precisely, we can arrange the elementary divisors into the following array

$$\begin{array}{ccc} p_1^{e_{1,1}} & \cdots & p_1^{e_{1,m}} \\ p_2^{e_{2,1}} & \cdots & p_2^{e_{2,m}} \\ \cdots & \cdots & \cdots \\ p_n^{e_{n,1}} & \cdots & p_n^{e_{n,m}} \end{array}$$

where $e_{i,j} = \begin{cases} e_{i,j}, & \text{if } j \leq k_i \\ 0, & \text{otherwise} \end{cases}$. Define $q_j = \prod_{i=1}^m p_i^{e_{i,j}}$, $j = 1, \cdots, m$. We also arrange the invariant subspaces into the following array

$$\begin{array}{ccc} C_{1,1} & \cdots & C_{1,m} \\ C_{2,1} & \cdots & C_{2,m} \\ \cdots & \cdots & \cdots \\ C_{n,1} & \cdots & C_{n,m} \end{array}$$

where $C_{i,j} = \begin{cases} C_{i,j}, & \text{if } j \leq k_i \\ \{0\}, & \text{otherwise} \end{cases}$. Define $D_j = \bigoplus_{i=1}^m C_{i,j}$, $j = 1, \cdots, m$.

Theorem 2. *(***Cyclic decomposition theorem of a finitely generated module over a principal ideal domain – in terms of invariant factors***) Suppose M is a finitely generated module over a principal ideal domain R, then*

$$M = D_1 \bigoplus \cdots \bigoplus D_m \bigoplus M_{free},$$

where M_{free} is a free submodule of M, and D_j is a cyclic submodule of M with order q_j, $j = 1, \cdots, m$. Moreover

$$q_m | q_{m-1}, q_{m-1} | q_{m-2}, \cdots, q_2 | q_1.$$

q_i, $i = 1, \cdots, m$ are called **invariant factors** *of M, which are uniquely determined by M up to the equivalence of associativity.*

19.15.3 Decomposition of a finite dimensional linear vector space under a linear operator

Now we come back to our original problem: *given a linear vector space V over a field K, for a given linear operator A on V, find a suitable set of basis for V under which the matrix representation of A has a nice form.*

In view of the module theory presented in the previous section, we note the polynomial ring $K[\lambda]$ is a principal ideal domain and V is a finitely generated torsion module over $K[\lambda]$ with the action defined by $p(\lambda)a = p(A)a$, $\forall p(\lambda) \in K[\lambda]$ and $a \in A$. $\mathrm{ann}(V) = \{p(\lambda) \in K[\lambda] : p(\lambda)V = \{0\}\}$ is an ideal of $K[\lambda]$, so there exists a unique $m(\lambda) \in K[\lambda]$ such that the coefficient of the term with highest degree is 1 and $\mathrm{ann}(V) = \langle m(\lambda) \rangle$. $m(\lambda)$ is called the **minimal polynomial** of A.

Theorem 3. *Suppose V is a finite dimensional linear vector space and A is a linear operator on V. Suppose the minimal polynomial $m(\lambda)$ of A has the decomposition into prime elements*

$$m(\lambda) = p_1^{e_1}(\lambda) \cdots p_n^{e_n}(\lambda),$$

where $p_i(\lambda)$, $i = 1, \cdots, n$ are non-associated monic order polynomial. Then V can be decomposed into direct sum

$$V = V_{p_1} \bigoplus \cdots \bigoplus V_{p_n},$$

where $V_{p_i} = \{v \in V : p_i^{e_i}(\lambda)v = 0\}$ and the minimal polynomial of $A|_{V_{p_i}}$ is $p_i^{e_i}(\lambda)$, $i = 1, \cdots, n$.

Each V_{p_i} $(i = 1, \cdots, n)$ can be further decomposed into the direct sum of cyclic subspace

$$V_{p_i} = \langle v_{i,1} \rangle \bigoplus \cdots \bigoplus \langle v_{i,k_i} \rangle$$

where $A|_{\langle v_{i,j} \rangle}$ has minimal polynomial $p_i^{e_{i,j}}(\lambda)$ $(j = 1, \cdots, k_i)$ and

$$e_i = e_{i,1} \geq e_{i,2} \geq \cdots \geq e_{i,k_i} \geq 1.$$

The elementary divisors $p_i^{e_{i,j}}(\lambda)$ of V are uniquely determined by A. In summary, we have

$$V = (\langle v_{1,1} \rangle \bigoplus \cdots \bigoplus \langle v_{1,k_1} \rangle) \bigoplus \cdots \bigoplus (\langle v_{n,1} \rangle \bigoplus \cdots \bigoplus \langle v_{n,k_n} \rangle).$$

If $\deg p_i^{e_{i,j}}(\lambda) = d_{i,j}$, then

$$\mathcal{B}_{i,j} = (v_{i,j}, Av_{i,j}, \cdots, A^{d_{i,j}-1} v_{i,j})$$

is a basis of $\langle v_{i,j} \rangle$. The matrix of A relative to the basis $\mathcal{B} = (\mathcal{B}_{1,1}, \cdots, \mathcal{B}_{n,k_n})$ is a block diagonal matrix

$$[A]_{\mathcal{B}} = \begin{bmatrix} C[p_1^{e_{1,1}}(\lambda)] & & & & & & \\ & \ddots & & & & & \\ & & C[p_1^{e_{1,k_1}}(\lambda)] & & & & \\ & & & \ddots & & & \\ & & & & C[p_n^{e_{n,1}}(\lambda)] & & \\ & & & & & \ddots & \\ & & & & & & C[p_n^{e_{n,k_n}}(\lambda)] \end{bmatrix}$$

where

$$C[p(x)] := \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

for $p(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$. *The block diagonal matrix* $[A]_{\mathcal{B}}$ *is called the* **rational canonical form** *of* A.

As a by-product, by using the rational canonical form of A, we can see the characteristic polynomial $|\lambda I - A|$ is equal to the product of elementary divisors $\prod_{i,j} p_i^{e_{i,j}}(\lambda)$.

When K is algebraically closed, the minimal polynomial $m(\lambda)$ of A can be decomposed into product of linear factors. This allows us to get a simpler form of the matrix representation of A.

Theorem 4. *If the minimal polynomial* $m(\lambda)$ *of* A *can be split on* K, *i.e.*

$$m(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_n)^{e_n},$$

then V *can be decomposed into*

$$V = (\langle v_{1,1}\rangle \bigoplus \cdots \bigoplus \langle v_{1,k_1}\rangle) \bigoplus \cdots \bigoplus (\langle v_{n,1}\rangle \bigoplus \cdots \bigoplus \langle v_{n,k_n}\rangle),$$

where $\langle v_{i,j}\rangle$ *is a cyclic subspace of* V *and the minimal polynomial of* $A|_{\langle v_{i,j}\rangle}$ *is* $(x-\lambda_i)^{e_{i,j}}$ *with*

$$e_i = e_{i,1} \geq e_{i,2} \geq \cdots \geq e_{i,k_i} \geq 1.$$

These elementary divisors are uniquely determined by A. *Let*

$$\mathcal{G}_{i,j} = (v_{i,j}, (A-\lambda_i)v_{i,j}, \cdots, (A-\lambda_i)^{e_{i,j}-1}v_{i,j}).$$

Then $\mathcal{G}_{i,j}$ *is a basis for* $\langle v_{i,j}\rangle$ *and the matrix representation of* A *under the basis* $\mathcal{G} = (\mathcal{G}_{1,1}, \cdots, \mathcal{G}_{n,k_n})$ *is the* **Jordan canonical form**

$$[A]_{\mathcal{B}} = \begin{bmatrix} g(\lambda_1, e_{1,1}) & & & & & & \\ & \ddots & & & & & \\ & & g(\lambda_1, e_{1,k_1}) & & & & \\ & & & \ddots & & & \\ & & & & g(\lambda_n, e_{n,1}) & & \\ & & & & & \ddots & \\ & & & & & & g(\lambda_n, e_{n,k_n}) \end{bmatrix},$$

where $g(\lambda_i, e_{i,j})$ *is the* $e_{i,j} \times e_{i,j}$ *matrix*

$$g(\lambda_i, e_{i,j}) = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & & \lambda_i \end{bmatrix}.$$

So far so good, but a fundamental question remains unanswered: *how do we find the elementary divisors so that we can write out directly the Jordan canonical form?* This is the content of the next section.

19.15.4 Computation of elementary divisors and invariant factors

This section is based on Qiu [11] (Reading Material 3, 5 and 7). Suppose R is a ring and we can define a matrix over R, whose entries are elements of R. Elementary row operations of matrices over R include swapping two rows, multiplying a row by an invertible element of R, and adding the multiple of one row to another row. Elementary column operations are defined similarly. By Cramer's rule, a matrix over R is invertible if and only if its determinant is an invertible element of R.

In particular, we consider the case of $R = \mathbb{C}[\lambda]$, the polynomial ring over the complex number field $\mathbb{C}$. Note the collection of invertible elements of $\mathbb{C}[\lambda]$ is just $\mathbb{C}$. For any square matrix A, we would like to find its elementary divisors.

Qiu [11] defined the invariant factors and elementary divisors for $A(\lambda) = \lambda I - A$. The invariant factors and elementary divisors thus defined are the same as those defined in Gong [3]. However, proofs that these differently defined concepts are the same cannot be found in both sources. I don't find a nice exposition elsewhere of the equivalence at this moment, so I'll only focus on the algorithmic aspect which allows us to calculate things explicitly.

Theorem 5. *Let A be a given square matrix over the complex field $\mathbb{C}$. Define $A(\lambda) = \lambda I - A$. Through the elementary row and column manipulation of $A(\lambda)$ as an element of the ring $\mathbb{C}[\lambda]$, $A(\lambda)$ can be reduced to a diagonal matrix. Each entry on the diagonal line can be factorized into the form of $(\lambda-\lambda_1)^{p_1}(\lambda-\lambda_2)^{p_2}\cdots(\lambda-\lambda_k)^{p_k}$, and each $(\lambda-\lambda_i)^{p_i}$ is an elementary divisor of $A(\lambda)$, corresponding to the $p_i \times p_i$ Jordan block*

$$\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}.$$

19.15.5 Examples

Example 1. $A = \begin{bmatrix} 2 & 3 & 2 \\ 1 & 8 & 2 \\ -2 & -14 & -3 \end{bmatrix}$. Its λ-matrix can be reduced to

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-3)^2 & 0 \\ 0 & 0 & \lambda-1 \end{bmatrix}.$$

So the elementary divisors are $(\lambda-3)^2$ and $(\lambda-1)$, and the minimal polynomial is $(\lambda-3)^2(\lambda-1)$. The Jordan canonical form of A is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix},$$

which can be verified by the **jordan** function of Matlab or the **JordanDecomposition** function of Mathematica. The invariant factors can be obtained by looking up the following array (see Theorem 2 or Qiu [11], Reading Material 7)

$$\begin{matrix} (\lambda-3)^2 & 1 & 1 \\ (\lambda-1) & 1 & 1 \end{matrix}$$

So $d_1(\lambda) = d_2(\lambda) = 1$ and $d_3(\lambda) = (\lambda-3)^2(\lambda-1)$.

Example 2. $A = \begin{bmatrix} 3 & -4 & 0 & 2 \\ 4 & -5 & -2 & 4 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 2 & -1 \end{bmatrix}$. Its λ-matrix can be reduced to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (\lambda+1)^2 & 0 \\ 0 & 0 & 0 & (\lambda-1)^2 \end{bmatrix}.$$

So the elementary divisors are $(\lambda+1)^2$ and $(\lambda-1)^2$, and the minimal polynomial is $(\lambda+1)^2(\lambda-1)^2$. The Jordan canonical form of A is

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

The invariant factors are $d_1(\lambda) = d_2(\lambda) = d_3(\lambda) = 1$ and $d_4(\lambda) = (\lambda+1)^2(\lambda-1)^2$, obtained by looking up the following array

$$\begin{matrix} (\lambda+2)^2 & 1 & 1 & 1 \\ (\lambda-1)^2 & 1 & 1 & 1 \end{matrix}$$

Example 3. $A = \begin{bmatrix} 3 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 3 & 0 & 5 & -3 \\ 4 & -1 & 3 & -1 \end{bmatrix}$. Its λ-matrix can be reduced to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & (\lambda-2)^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & (\lambda-2)^2 \end{bmatrix}.$$

So the elementary divisors are $(\lambda-2)^2$ and $(\lambda-2)^2$, and the minimal polynomial is $(\lambda-2)^2$. The Jordan canonical form of A is

$$\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

The invariant factors are $d_1(\lambda) = d_2(\lambda) = 1$, $d_3(\lambda) = d_4(\lambda) = (\lambda-2)^2$, obtained by looking up the following array

$$\begin{matrix} (\lambda-2)^2 & (\lambda-2)^2 & 1 & 1 \end{matrix}$$

19.16 Numerical Range

1.

Proof. By Theorem 8 of Chapter 8, we can find an orthonormal basis consisting of eigenvectors of A. Let $a_1, \cdots, a_n$ be the eigenvalues of A (multiplicity counted) with $v_1, \cdots, v_n$ the corresponding eigenvectors. For any $x \in X$, we can find $\theta_1(x), \cdots, \theta_n(x) \in \mathbb{C}$ such that $x = \sum_{i=1}^n \theta_i(x) v_i$. Then

$$|(Ax, x)| = \left| \sum_{i,j} \theta_i(x)\bar{\theta}_j(x)(Av_i, v_j) \right| = \left| \sum_{i,j} \theta_i(x)\bar{\theta}_j(x)(a_i v_i, v_j) \right| = \left| \sum_{i=1}^n a_i |\theta_i(x)|^2 \right| \leq \max_{1 \leq i \leq n} |a_i| = r(A).$$

Combined with (2), we conclude $r(A) = w(A)$. □

2.

Proof. By definition $\|A\| = \sup_{\|x\|=1} \|Ax\|$. Using the notation in the solution to Exercise 1, we have

$$Ax = \sum_i \theta_i(x) A v_i = \sum_i \theta_i(x) a_i v_i.$$

So $\|Ax\| = \sqrt{\sum_{i=1}^n |a_i|^2 |\theta_i(x)|^2} \le r(A) = w(A)$, where the last equality comes from Exercise 1. This implies $\|A\| \le w(A)$. By Theorem 13 (ii) of Chapter 7, $w(A) \le \|A\|$. Combined, we conclude $w(A) = \|A\|$. □

3.

Proof. To verify (7), we note

$$\prod_k (1 - r_k z) = \prod_k (\bar{r}_k - z) \cdot \prod_k r_k = \prod_k (\bar{r}_k - z) \cdot e^{\frac{2\pi i}{n} \cdot \sum_{k=1}^n k} = \prod_k (\bar{r}_k - z) \cdot e^{(n+1)\pi i} = (-1)^{n+1} \prod_k (\bar{r}_k - z).$$

Since $(\bar{r}_k)^n = 1$, $\bar{r}_1, \cdots, \bar{r}_n$ are a permutation of $r_1, \cdots, r_n$. So

$$\prod_k (\bar{r}_k - z) = (-1)^n (z^n - 1).$$

Combined, we conclude $(1 - z^n) = \prod_k (1 - r_k z)$. To verify (8), we use (7) to get

$$\frac{1}{n} \sum_j \prod_{k \neq j} (1 - r_k z) = \frac{1}{n} \sum_j \frac{1 - z^n}{1 - r_j z}.$$

$\sum_j \frac{1}{1 - r_j z}$ is a rational function over the complex plane $\mathbb{C}$, which can be assumed to have the form $\frac{P(z)}{Q(z)}$ with $P(z)$ and $Q(z)$ being polynomials without common factors. Since $r_1, \cdots, r_n$ are singularity of degree 1 for $\sum_j \frac{1}{1 - r_j z}$, we conclude $Q(z) = \prod_k (1 - r_k z) = 1 - z^n$, up to the difference of a constant factor. Since $\sum_j \frac{1}{1 - r_j z}$ has no zeros on complex plane, we conclude $P(z)$ must be a constant. Combined, we conclude

$$\sum_j \frac{1}{1 - r_j z} = \frac{C}{1 - z^n}$$

for some constant C. By letting $z \to 0$, we can see $C = n$. This finishes the verification of (8). □

4.

Proof. If $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $(Ax, x) = x_1^2 + x_2^2 + x_1 x_2$. If $x_1^2 + x_2^2 = 1$, we have $(Ax, x) = 1 + x_1 \cdot \mathrm{sign}(x_2)\sqrt{1 - x_1^2}$. Calculus shows $f(\xi) = \xi\sqrt{1 - \xi^2}$ $(-1 \le \xi \le 1)$ achieves maximum at $\xi_0 = \frac{\sqrt{2}}{2}$. So $w(A) = 1 + \frac{1}{2} = \frac{3}{2}$. Similarly, plain calculation shows $w(A^2) = 2$. □

References

[1] M. Engeli, T. Ginsburg, H. Rutishauser and E. Stiefel. *Refined iterative methods for computation of the solution and the eigenvalues of self-adjoint boundary value problems*, Birkhauser Verlag, Basel/Stuttgart, 1959.

[2] Gene H. Golub and Charles F. van Loan. *Matrix computation*, 3rd Edition. Johns Hopkins University Press, 1996.

[3] Gong Sheng. *Five lectures in linear algebra* (in Chinese), Science Press, Beijing, 2004.

[4] Gong Sheng and Gong Youhong. *Concise complex analysis*, Revised Edition, World Scientific, 2007.

[5] J. Munkres. *Analysis on manifolds*, Westview Press, 1997.

[6] P. Lax. *Functional analysis*, Wiley-Interscience, 2002.

[7] Steven Huss-Lederman, Elaine M. Jacobson, Anna Tsao, Thomas Turnbull, Jeremy R. Johnson. Implementation of Strassen's algorithm for matrix multiplication. Proceedings of the 1996 ACM/IEEE conference on Supercomputing (CDROM), p.32-es, January 01-01, 1996, Pittsburgh, Pennsylvania, United States.

[8] Nie Ling-Zao and Ding Shi-Sun. *Introduction to algebra* (in Chinese), 2nd Edition, Higher Education Press, 2000.

[9] M. Papadrakakis. A family of methods with three-term recursion formulae. *International Journal for Numerical Methods in Engineering*, Vol. 18, 1785-1799 (1982).

[10] Qiu Wei-Sheng. *Advanced algebra* (in Chinese), Volume 1, Higher Education Press, 1996.

[11] Qiu Wei-Sheng. *Advanced algebra* (in Chinese), Volume 2, Higher Education Press, 1996.

[12] Kosaku Yosida. *Functional analysis*, 6th Edition. Springer, 1996.